WPS4377

POLICY RESEARCH WORKING PAPER 4377

Early Identification of At-Risk Youth in Latin America:

An Application of Cluster Analysis

Emilie Bagby
Wendy Cunningham

The World Bank
Latin America and the Caribbean Region
Human Development Department
October 2007

Policy Research Working Paper 4377

Abstract

A new literature on the nature of and policies for youth in Latin America is emerging, but there is still very little known about who are the most vulnerable young people. This paper aims to characterize the heterogeneity in the youth population and identify ex ante the youth that are at-risk and should be targeted with prevention programs. Using non-parametric methodologies and specialized youth surveys from Mexico and Chile, the authors quantify and characterize the different sub-groups of youth, according to the amount of risk in their lives, and find that approximately 20 percent of 18 to 24 year old Chileans and 40 percent of the same age cohort in Mexico are suffering the consequences of a range of negative behaviors. Another 8 to 20 percent demonstrate factors in their lives that pre-dispose them to becoming at-risk youth – they are the candidates for prevention programs. The analysis finds two observable variables that can be used to identify which children have a higher probability of becoming troubled youth: poverty and residing in rural areas. The analysis also finds that risky behaviors increase with age and differ by gender, thereby highlighting the need for program and policy differentiation along these two demographic dimensions.

This paper—a product of the Latin American and Caribbean Region , Human Development Department—is part of a larger effort in the department to understand the challenges facing at-risk youth in the region and to design effective policy to support them. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at Wcunningham@worldbank.org.

EARLY IDENTIFICATION OF AT-RISK YOUTH IN LATIN AMERICA: AN APPLICATION OF CLUSTER ANALYSIS

Emilie Bagby and Wendy Cunningham[1]

[1] University of Illinois Urbana-Champaign and World Bank, respectively.

I. Introduction

The specific needs of young people are an increasingly debated subject in the development community. A large literature has emerged to inform the debate in Latin America, principally by mapping youth behaviors, positing policy and program interventions to address the behaviors, and, to a lesser extent, to identifying the (economic) factors driving young people's decisions (Lloyd 2005, World Bank 2006). A primary shortcoming of the research is that it treats youth as a homogenous group and reports average behaviors, thereby not capturing the complexities of the youth population across its many dimensions.

Even if we do understand the heterogeneity in the youth population, appropriate policy requires an understanding of why young people make the decisions they do. The evidence for Latin American youth is limited. The role of economic incentives and budget constraints affecting decision-making by young people in developing countries has been examined (World Bank 2006 and the sources within), building on Gruber's (2001) work in the United States. A second line of work considers the broader context in which young people form their preferences and make their decisions (World Bank 2003, 2007), building on the public health and psychology research in the United States. While this literature gives good insights as to *how* to modify behaviors, it is silent on *whom* prevention programs should target.

This paper aims to construct a more complete picture of the heterogeneous youth population in LAC by quantifying and describing the at-risk youth population and by identifying easily observable variables can be used to identify, ex ante, those youth who have a higher propensity of engaging in risky behaviors. We use non-parametric methodologies and special cross sectional youth surveys from Mexico and Chile that permit us to identify the factors that today's risk-taking youth had in their childhoods, which are then used to identify who should be the target of prevention policies and programs.[2]

The paper finds that over 20 percent of 18 to 24 year olds in Chile and 40 percent in Mexico have influences in their early and current lives that predispose them to negative behaviors, they have engaged in these behaviors, and they are suffering the consequences. Conversely, 40 percent of Chileans age 18 to 24 and 16 percent of Mexicans of the same age do not display any of these risks, behaviors, or consequences. While many of the factors that identify if a person will engage in risky behavior are unobservable, we find two observable variables that are highly correlated with these underlying factors and can be used

[2] Causality can only be determined using panel data that includes information on behaviors and household factors, which do not exist in LAC, to the best of our knowledge.

to identify which children have a higher probability of becoming troubled youth: poverty and residing in rural areas. While self-identifying as indigenous is correlated with risky behaviors, the relationship is not statistically significant. And, we find that risky behaviors increase with age and differ by gender, thereby highlighting the need for program and policy differentiation along these two dimensions.

There are five sections following this introduction. Section II presents the conceptual framework and provides a brief review of the literature. Section III describes the methodology and Section IV discusses the data. The results are presented in section V and section VI concludes.

II. Review of the Literature and Conceptual Framework

The issue of youth development is relatively new in the field of economics. Perhaps the most extensive study in the economic literature is Gruber (2001), which investigates the determinants and implications of nine different behaviors among US youth – smoking, risky driving, sexual activity, suicide, marijuana use, crime, alcohol use, school dropout, and mis-nutrition – using both cross sectional and time series data. The rational addiction model developed by Becker and Murphy (1988) is expanded to allow for youth-specific characteristics identified in the psychology and human development literature, namely myopia, time inconsistent preferences, and projection bias.[3] The book concludes that youth respond to incentives such as age-specific legal penalties, prices, and income and that the marginal cost to additional risk taking is small once participation in risk behavior has begun. Recent studies have tested these conclusions in the context of developing countries, finding similarities and differences with the US[4].

A shortcoming of the economic literature is the absence of a discussion about preference formation and that early experiences vary across the youth population, leading to heterogeneity in decision-making during the youth years.[5] In contrast, an extensive literature in the public health and psychology fields start from the assumption that preference formation and constraints before the youth years, as well as during them, explain a significant portion of the variation in youth behaviors. The ecological risk framework posits that youth are a product of individual (personal), micro-, and macro-environmental factors (Bronfenbrenner 1979). The individual factors are those skills, behaviors, and ideas that are

[3] "Projection bias" is understood as today's preferences may not be representative of future adult preferences.
[4] See Lloyd ed. (2005), World Bank (2006), and Attanasio et al. (2005) for example.
[5] Gruber (2001) discusses preferences, but it focuses on the time-inconsistency of preferences between the youth period and adulthood. It does not investigate why different youth have different preferences.

"hardwired", rather than formed, such as rage, optimism, or general health. The micro factors include preferences taught and formed by the family, peers, community, and local institutions and the constraints imposed by the same, including household poverty. The macro factors include more general influences and constraints, such as gender/race discrimination, armed conflict, poverty and economic inequality. These factors are commonly classified into two groups: the set of personal, macro, and micro factors that increase the risk of negative behaviors (risk factors) and the set that prevent youth from engaging in negative behaviors, commonly called protective factors. Each person has a set of risk and protective factors that influence preference formation, constraints, and thus behaviors.

Most of the empirical testing of the ecological risk framework depends on correlations between current risky behaviors and personal, micro, and macro factors of the youth population. The literature in the US primarily focuses on identifying factors related to single risky behaviors (e.g. substance use).[6] Recent US literature has begun to take advantage of longitudinal data available in the US[7] to demonstrate that many of those factors correlated with risk taking behaviors are actually causal factors.

The few studies that have tested the model using data from Latin American and Caribbean find similar results to those in the US. Blum (2004) uses data collected in 11 Caribbean countries and finds that a positive relationship with a caring adult, whether in the family or in school, is a key factor that is positively correlated with less risky sexual behavior (sexual debut and condom use), contraception, pregnancy and childbearing. Lloyd (2005) identifies the positive effects of schooling and health on transitions to risk-free adulthood

[6] There is an extensive literature in the US which is too large to cover here. The US National Library of Medicine and National Institute of Health maintains a web page (http://www.ncbi.nlm.nih.gov/sites/entrez), that lists hundreds of published articles that have used this methodology. A few of the papers that motivated this paper include: Resnick et al. (2004) look at risk and protective factors related to youth violence; Blum et al. (2002) describe the ecological risk framework and provide empirical evidence for three risky behaviors (weapon related violence, ever had sexual intercourse and ever used cocaine); Scal et al. (2003) look at risk and protective factors related to smoking; Zweig et al (2002) identify methods of predicting risk profiles using risk and protective factors such as psychosocial development, school and family characteristics using OLS and multinomial logit regressions with cross sectional data; and Bernat and Resnick (2006) provides a comprehensive review of the resiliency framework and additional empirical support for this framework in promoting healthy youth development.

[7] These studies, which also appear on the NIH web site, use the National Longitudinal Survey of Adolescent Health (ADDHEALTH) through the University of North Carolina, Carolina Population Center, surveys youth in grades 7-12 with the first wave of interviews in 1994 (http://www.cpc.unc.edu/projects/addhealth/design). Follow-up waves were in 1996 and 2002, the latter enabling more detailed analysis. This survey is school based and asks about risk and protective factors as well as behaviors. The National Longitudinal Survey of Youth (NLSY) 1997 from the Bureau of Labor Statistics, surveyed males and females born in 1980-1984 (http://www.bls.gov/nls/nlsy97.htm) and focuses primarily on educational and employment outcomes. The previous NLSY was from 1979.

using DHS surveys from developing countries. World Bank (2007) finds that low self-esteem, spirituality, school connectedness, abuse in the home, abuse in the community, connectedness to institutions, poverty and gender are the factors most associated with risky behaviors and negative outcomes in Brazil.[8] Youth in the English-speaking Caribbean who have lower risky behaviors[9] are also those who are more connected to parent/family, attending religious services, feel little rage, have not been abused or witnessed parental violence, do not have family members who have attempted suicide, have mentally healthy parents, and have households free of illegal substances (World Bank 2003). A series of papers by Brook, et. al. (2001, 2002a, 2002b) find similar correlations between household factors, poverty, and community factors for explaining substance use and violence among Colombian youth, while Hutz and Silva (2003) find that young Brazilian men who have been incarcerated for violence are disproportionately the sons of poor, uneducated, and violent fathers.

The ecological risk framework can be stated more formally. A person *i* has a set of behaviors, B_i, that are determined by a vector of risk factors, r_i, and protective factors, p_i, determined at the individual (I_i), micro (c_i), and macro (M_i) levels.

$$B_i = f\left(r\left(I_i, c_i, M_i\right), p\left(I_i, c_i, M_i\right)\right) \qquad (1)$$

If an element in any of the vectors I_i, c_i, or M_i leads to a positive behavior in B_i, it will take a positive value in $p\left(I_i, c_i, M_i\right)$ and a 0 in $r\left(I_i, c_i, M_i\right)$. Likewise, an element that leads to risky behavior in B_i will take a 0 value in $p\left(I_i, c_i, M_i\right)$ and a positive value in $r\left(I_i, c_i, M_i\right)$. A weighted average of the risk (r_i) and protective factors (p_i) specific to each person will predict the behavior elements in the vector B_i. Behaviors include elements such as unprotected sex, school truancy, or substance use.

The outcomes of these behaviors are a function of the behaviors, the individual, micro, and macro environments, and luck (δ). The outcome, O_i, is given by

$$O_i = f\left(B_i, I_i, c_i, M_i, \delta\right) \qquad (2)$$

[8] Behaviors and outcomes studied include: grade repetition, early labor force entry, early sexual initiation, risk taking sexual practices, alcohol use tobacco use, illegal drug use, violence and suicide attempt.
[9] Behaviors and outcomes studied include: perception of general health, ever had sexual intercourse, ever attempted suicide, violent behavior, problems due to alcohol and drugs.

Outcomes can be good (school completion, youth participation) or negative (school dropout, exclusion). Risk and protective factors are included in the function since they can magnify or mitigate the outcomes of a behavior. We assume a distribution of δ that is constant across people, but instrumental in determining if behavior B_i becomes outcome O_i.

Equations 1 and 2 can be used to link the concepts of risk/protective factors, behaviors, and outcomes to levels of risk and give insight to policy. A person with high values of the elements in $r(I_i, c_i, M_i)$, low values of the elements in $p(I_i, c_i, M_i)$, and a B_i and O_i that displays few negative behaviors and negative outcomes, is defined as being type I risk. More generally, a person classified as type I risk has many risk factors and few protective factors, indicating that they have a predisposition to engage in negative behaviors, but the person has not undertaken any risky behaviors. Prevention programs would be targeted to this group.[10] A person with high values of the individual elements in $r(I_i, c_i, M_i)$, low values of the individual elements in $p(I_i, c_i, M_i)$, a B_i that displays many negative behaviors, and an O_i, with few negative outcomes would be classified as type II risk, where risk factors are present, protective factors are few, and the young person has engaged in risk-taking behaviors without having experienced any negative consequences. These youth are at-risk of suffering consequences and can thus benefit from prevention or second chance programs. Youth with positive values in the O_i vector are those who are suffering the consequences of their behaviors and are thus categorized as type III risk. These youth tend to have high risk factors, poor protective factors, and many negative behaviors, many of which might be identified prior to their suffering the consequences of their risky behaviors. They are candidates for second chance (remedial or rehabilitation) programs.

III. Methodology

To understand heterogeneity in the youth population, we use cluster analysis to identify different groups of youth based on the observable elements in the vectors I_i, c_i, M_i, B_i, and O_i. Classifying these groups along the risk profiles will allow us to measure the share of the youth population that is at different levels of risk and to identify observable

[10] See Blum (1998) for a discussion of resiliency based intervention programs.

factors that can be used by policymakers to target programs to youth with type I, II, or III risk.

A. Cluster Analysis

Cluster analysis is a means to identify correlations across large data sets without imposing, a priori, a structure on the data. Observations are grouped based on minimizing a distance measure between each variable for each observation, i.e. the observations in a cluster share a set of common variables. By comparing the mean values of various variables across clusters, we can characterize each cluster.

Ward's method (minimum-variance) of testing was selected since it provided the most distinct and interpretable clusters.[11] Ward's method uses the error sum of squares criteria.[12] The variance is minimized by calculating the sum of squared errors from the mean of the cluster for each of the m variables for each observation:

$$W = \sum_{k}\sum_{j}\sum_{i}(x_{ijk} - \bar{x}_{jk})^2 \qquad (3)$$

$i = 1,...., n$ observations, $j = 1,..., m$ variables, and $k = 1,..., l$ clusters

Initially, each of the n observations forms its own cluster. The first merge is identified by calculating the sum of squares for each pair of cluster. The pairing with the smallest sum of squares is identified and those clusters are joined, leaving n-1 clusters. The second grouping calculates the sum of squared errors again and pairs the two clusters that have the smallest value, leaving n-2 clusters. The process is repeated until the optimal number of clusters has been reached.

Three tools were used to determine the optimal number of clusters. First, stop commands following two possible rules (Calinsky-Harabasz and Duda-Hart) were used to find criterion for each cluster possibility.[13] The Calinski and Harabasz method suggests the optimal number of clusters (g) that maximizes an index C(g) which uses the pooled within-cluster covariance matrix (W) and the between-cluster covariance matrix (V), where

$$C(g) = [\text{trace } (V)/(g-1)]/[\text{trace } (W)/(n-g)] \qquad (4)$$

The Duda and Hart method maximizes D(g), where

[11] There are many different ways to perform cluster analysis, and no particular method is considered the best. Ward's linkage cluster analysis is a commonly used agglomerative hierarchical method.

[12] An attractive feature of the Ward's method is that it performs well with groups that are of unequal size, which, as will be shown in the results, strongly characterizes the data. See Everitt et al. (2001) for a theoretical discussion of Cluster Analysis and Ward's criterion. See Cunningham and Maloney (2001) for an application.

[13] These two methods are implemented in STATA. They were identified as the two best methods available (out of 30) by Milligan and Cooper (1985) and are discussed in Everitt et al. (2001).

$$D(g) = Je(2)/Je(1) \quad (5)$$

Je(2) is the sum of the within cluster sum of squared distances between the objects and centroid if the cluster is split into two and Je(1) is the within cluster sum of squared distances. The local criteria calculated in equations 4 and 5, C(g) and D(g), are then combined with test statistics for each clustering option to suggest the optimal number of clusters. Larger values of both methods indicate that the clusters are more distinct from each other while lower values indicate that the clusters are not very different from each other and therefore artificially sub-divided.

Second, dendrograms were used to select among the multiple "right" clusters that the other methods may give. Dendrograms graphically depict the hierarchical relationship between the clusters by showing the order in which clusters are merged as well as the distance between the clusters. At each level of the cluster formation process, a dendrogram can be generated to view the relationships between the clusters. The dendrogram changes as clusters are grouped and un-grouped, thus enabling the researcher to optimally choose the clustering level.

Third, once the optimal number of clusters is suggested and the dendrograms generated, the clusters themselves are observed and the means of the variables are compared across clusters. The differences found between the different clusters are used to ultimately determine the optimal number of clusters. For instance, if 5 clusters were suggested, 6 clusters were investigated to see if there was an interpretable difference. If not, 5 clusters were used, however if so, 6 clusters were used. This process could then repeat.

The cluster analysis is performed for 10 cohorts, identified by age, gender, and country. The sample for each country is divided ex-ante by age and sex since, when pooling the sample, these two variables dominated the clusters to such an extent that the risk and protective factors of interest played a very small role. Six cohorts from Mexico are analyzed: female ages 12 to 14, female ages 15 to 17, female ages 18 to 24, and males in each of the three age groups. Only the four older cohorts from Chile are analyzed since youth age 12 to 14 were not included in the sample.

While the objective of a cluster analysis is to identify which variables move together, a decision was made to treat some variables endogenously and others exogenously. For example, a hypothesis is that poverty status is a good indicator for a youth being "at-risk". If we use this variable to create the clusters, it is possible that poverty is such a strong factor that it drives the clusters and renders the other variables meaningless. Thus, for these

type variables, we carried out the analysis treating them as both endogenous and exogenous and found little difference. We thus report only the results for treating them exogenously.

The advantage of cluster analysis is that the only priors required are in the variable construction, such that they range between 0 and 1. For continuous variables, the value was normalized:

$$x = \frac{y}{y_{\max}} \tag{6}$$

Binary variables were assigned a 0 or 1 and discrete variables were assigned a value between 0 and 1 based on the ordering of the responses. A variable takes a value of 1 the closer it is to the variable being described. For example, the variable "abuse" takes a value of 1 if there is abuse in the household and a 0 if there is not while the variable "connected" takes a value of 1 if the respondent reports that they reach out for help all of the time if they have problems, a value of 0 if they never reach out, and a value in between depending on how frequently they reach out. Variables that could not be ordered in a logical way were not included in the analysis.

B. Identification of Correlation between Factors and Behaviors

Using the generated clusters, we use parametric methods to test the relationship between behaviors/outcomes and factors. First, we graph the clusters using bubble graphs to show the relationships between risk and protective factors with positive behaviors[14] then we regress the average level of positive behaviors from each cluster against the average level of risk factors and protective factors from each cluster, [15] using OLS to identify if there is a significant relationship between the behavior and the risk or protective factors. We consider the relationships in both countries for each age cohort and for the sample as a whole. We also test the correlation between positive behaviors and three easily observable variables that seem correlated with good behavior: low poverty, ethnic majority, and urban residence.

[14] The size of the bubble represents the sample size for each cluster within the age/sex groups.

[15] For this exercise, we need to assign a "good" or "bad" subjective ranking to each I, c, M, B, and O. This is straightforward for most cases, but in three instances, the variable changed value from 0 to 1 and vice versa depending on whether the youth is younger than 18 years of age or not: parenthood, working, and marriage. The case can be made that for those under 18, it is not optimal to be a parent, to be working, or to be married. Indeed, this opinion is supported by the large body of research attempting to explain how and why youth arrive at these adult outcomes at an early age. At the age of 18, however, it is inappropriate to consider a person with these characteristics to be necessarily at-risk. Therefore, for those under 18 years of age, the indicators for being a parent, being married and working take a 0, whereas for those 18 to 24 the indicators take a value of 1.

IV. Data Description

A. Data

As youth departments and governments become more sophisticated in their efforts to understand youth, several LAC countries have developed specialized youth surveys. We use two surveys in this paper: the 2003 National Youth Survey (*Encuesta Nacional de Juventud*) from Chile and the 2000 National Youth Survey from Mexico. Youth are defined as being between the ages of 12 and 24 for this analysis, consistent with World Bank (2006).

Chile first implemented their National Youth Survey in 1994 and has subsequently repeated it every three years. This paper uses the most recent data set that could be accessed, which was collected in 2003. A nationally representative sample of youth aged 15 to 29 is surveyed, reaching all regions (urban and rural) in the country, with a sampling error of 2.1% and a 95% confidence level. The sample size for youth aged 15 to 24 used in the analysis is 5321[16].

The survey questions cover socioeconomic status, youth behaviors and opinions, and family backgrounds. It includes many questions about risk factors such as family cohesion, neighborhood violence, and social exclusion. Many protective factors are not covered, but there is information about trust in institutions, connectedness, and relationship with parents. The behaviors and outcomes are limited to employment, schooling, sexual health, victimization, discrimination and participation in activities. It does not include questions about whether the respondents take drugs or commit violent acts. There are very limited health and health access data.

Mexico carried out its National Youth Survey in 2000, through the National Institute for Statistics, Geography, and Information (*Instituto Nacional de Estadística, Geografía e Informática*, INEGI).[17] The survey sample is nationally representative and was performed in two stages: the entire household was surveyed and asked basic household characteristics, and, at a later date, youth aged 12 to 29 were asked a set of youth-specific questions. Only those aged 12 to 24 were included in this analysis, resulting in a sample of 37,979 respondents[18].

---

[16] For this analysis, 195 observations out of 7,189 (2.7% of sample) for youth aged 15 to 29 were dropped due to missing data. We then restricted the sample to youth aged 15 to 24 dropping another 1674 observations. In many cases, missing responses could be coded based on responses to related questions so as to maintain a larger sample size. For instance, if someone does not respond as to whether they attend church, after they have already indicated that they do not believe in God, then we assume that they do not attend church. It would be useful to do analysis using censored data tools.

[17] The survey was repeated in 2005, but the data was not available at the time this paper was under preparation.

[18] Of the almost 60,000 youth respondents, about 10,000 youth aged 12-29 were not surveyed the second time, and were dropped from the sample used for the analysis. The reasons for not interviewing these youth were

These data are similar to those from Chile; however they are slightly stronger on protective factors and more limited on risk factors. The Mexican dataset does not have information about connectedness with other adults, abuse in home, victimization or violence in community, but it does have questions on attitudes towards drugs, the number of parents in the household and parental response to good/bad behavior. As with the Chile survey, there is information on individual earnings and three family poverty variables – parental earnings, parental education and durable goods in household – as well as regional/community level information. Although youth aged 12 to 29 were surveyed, those aged 12 to 14 were not asked questions about sexual activity or practices.

To the extent possible, similar variables are used in our analysis for both countries, although the Chilean data are more robust with the protective factors and the Mexican data include more risk factors. Table 1 describes each variable and identifies which variables are used in which clustering exercise and which are treated exogenously.

B. Descriptive Statistics

Descriptive statistics for continuous and binary variables are provided for risky behaviors and factors in Table 2.[19] Sample means for all of the normalized variables are presented in Annex A to provide a baseline of the average status of the population against which to compare the variable values in each distinct cluster. The Chileans are 1.7 years older in the sample than the Mexicans, but this is primarily due to the inclusion of 12 to 14 years olds in the Mexican survey. Thus, it is important to compare characteristics by age group, as presented in Table 2. Both surveys are slightly more weighted toward women. About 11 percent of the Chilean sample self-identifies as indigenous; ethnicity was not

tracked in the dataset: they did not want to participate, were not at home at the time of the interview and would not return within the week, were on vacation, were working or at school in another city, were disabled, and other. Comparing the poverty variables (education level and monthly earnings of heads of households) and rural means of this dataset before and after dropping the data showed no significant difference at α=.01 (However, at α=.05, the difference between the education level of heads of households the original and remaining datasets is not zero). An additional 5% of the observations were dropped in creation of the variables. Since this was such a small amount of the sample, the resulting dataset was not statistically different from the original. Finally, the data were restricted to youth aged 12 to 24, thus further decreasing the sample size by 8903.

[19] Variables with categorical responses cannot be summarized, but the mean of the variables created from the categorical responses can be. Thus, the categories are not summarized in Table 2, but the mean of the variable created from the categorical responses is summarized in Annex A.

identified in the Mexican survey. In Chile, about 13% of the sample is considered rural[20] while in Mexico about 24% is rural.

In both countries, a significant percentage of youth in both countries are dropping out of school early and not working after age 18. In the survey, 29.9% of Mexican and 14.2% Chilean youth have dropped out of school prior to completing high school, but 41% of Chileans age 18 to 24 and 27.5% of Mexicans in the same age range are in school.[21] While 20 to 30% of Mexican and Chilean youth are inactive after age 18, 12.8% of Mexican and 6.7% Chilean youth are inactive before age 18.

Chileans have earlier and riskier sex than Mexicans. While only 13% of Mexicans age 15 to 17 report having had their first sexual experience, 27% of Chileans in this age group report the same. Chileans are 10 percentage points less likely to use condoms than are Mexicans, which reflects that 2.5 percent of Mexicans and nearly 4 percent of Chileans age 15-17 have had their first child. Marriage rates in Chile are less than one-third of those in Mexico for each age group.[22]

In terms of the risk and protective factors at the micro-level, Mexican and Chilean youth report a high level of personal connections with someone. Mexican youth report a wide range of topics that they communicate with their parents about, from school and work to politics and religion. While Chileans report good relationships with their parents, ten percent of Chileans also note important relationships with other adults. On the negative side, six percent of Chileans have suffered abuse in the home and 8.5 percent report substance abuse in their homes.

Mexican youth tend to trust government institutions less than they trust local institutions. They feel that school quality is good on average, although half say they do not have access to health services. There is a connection to religious institutions as 21 percent of Chilean youth and 9 percent of Mexican youth attend church weekly, with 66 percent of Mexican youth attending church at least once in the past month.

Of the questions on individual risk/protective factors, more than 85 percent of Chilean youth are optimistic about the future, similar to rates in Brazil (Instituto Cidadania

[20] The sampling was done in communities with at least 2000 inhabitants, so a rural indicator means that the respondent comes from a community with between 2000 and 5000 inhabitants. An urban respondent lives in a community of at least 5000 people.

[21] If there is no repetition, youth would finish high school at age 17 or 18 in both countries.

[22] In Chile and Mexico, the legal age of marriage age is 18, but Mexican youth can marry as early as age 16 with parental consent.

2004), and an even greater share, 98%, of Mexicans have a sense of well-being. However, half of Mexican youth and one-fifth of Chilean youth feel socially excluded.

There are few questions on macro factors, but the data do show that over half of the Mexican youth are from families whose parents have less than a primary education, while over one-third of Chilean youth have parents with less than a primary education. In addition, over half of Chilean youth have felt discriminated against at some time.

V. Results

The clustering methodology identified four to seven sub-groups within each of the 10 age-sex-country cohorts, resulting in 53 (23 from Chile, 30 from Mexico) distinct groups of youth. To simplify the discussion, the commonalities across the clusters are presented in this section. The results are presented by risk category to allow us to characterize the heterogeneity in the youth population and to create a picture of youth in each category, based on the findings from the 53 clusters. Tables 3a and 3b present a summary of youth in the different risk categories in Mexico and Chile respectively while the full cluster results are presented in Tables 4a-4j. Annex B gives a brief description of each of the 53 clusters.

A. Who Are the Youth at Risk in Mexico and Chile?

Type III Risk

Thirty-three percent of Mexican youth can be classified as type III risk, meaning that they are coping with the outcomes, or consequences, of their risky behavior, while 16.8% of Chilean youth fall into this risk category (Tables 3a and 3b). Considering separately the age-gender categories better characterizes type III risk youth in Mexico and Chile.

Fourteen percent of Mexican youth aged 12 to 14 are in this risk category. They have almost all dropped out of school, resulting in an average education level of primary school. Few are involved in organized activities and a large proportion is working (49% of boys and 27% of girls). We cannot report on the sexual activity of these youth since they were not asked these types of questions. All youth who are illiterate are in this category (Tables 4a and 4b, column 1).

Eight percent of Chileans age 15 to 17 and almost 40 percent of Mexicans in the same age range are classified as risk type III (Tables 3a and 3b). Almost all have dropped out of school prior to completing secondary education, many are inactive, most engage in risky sexual behavior and feel socially excluded. Among the women, a strong at-risk mother

group emerges, comprised of women who engaged in risky sexual behavior at a young age, left school, had their child, and largely remain unmarried.[23] The males report unsafe sexual activity, low education level and idleness. They also report an early working age as well as an early age at onset of sexual activity (Tables 4c, 4g, 4h column 1; Table 4d columns 1-3).

Almost half of Mexican youth (43%) and almost one-quarter of Chilean youth (22%) aged 18 to 24 can be classified as risk type III. These young people have had significantly more time to engage in risky behaviors and thus demonstrate a wider range of negative behaviors and outcomes. They break into two general categories: inactive youth and parents. The inactive males dropped out of school early and are not working – these are the stereotypical youth at-risk. Inactive females tend to be mothers who dropped out of school early and continue to engage in unsafe sexual activities. Among male parents, the fathers that are working typically dropped out of school prior to completing secondary education. Some but not all of these youth engage in risky sexual practices (Tables 4e, 4f, 4i, 4j).[24]

Several of the risk factors clustered with the groups described above, and protective factors were notably absent. The 12-14 year old Mexicans in this category come from the poorest households, tend to live in more rural areas, come from abusive households, had little parental control/influence on their behavior related to alcohol and smoking, had low access to healthcare and high social exclusion. The youth in these clusters report a low to average level of spirituality and church attendance (Tables 4a and 4b).

Risk factors for the 15 to 17 year olds include low parental education, low level of connectedness, and non-positive family backgrounds. Thirty-three percent of the males in the Mexico type III risk clusters are from rural areas, but the Chilean type III clusters are not as rural with 25 to 27 percent living in rural areas[25]. Since Chile is much less rural overall than Mexico (13% compared to 25%), we can conclude that living in a rural area in Chile is more risky compared to living in a rural area in Mexico (Tables 4c, 4d, 4g, 4h).

The 18 to 24 year olds in this age category come from disadvantaged backgrounds, with parents having below a secondary education on average. Twenty-three percent of the Chilean women in this category and age group live in rural areas, whereas about 19% of

[23] While the girls and boys in this cluster have similar situations, the data do not well capture fatherhood, so the most risky boy clusters will not report children, pregnancy or marriage, where the girls do.
[24] The males age 18-24 clusters that are classified as risk type III are largely defined by school and work variables, while the women's clusters are also defined by sexual behavior. This is largely driven by the data in two ways. First, the data about risky behavior sexual is largely targeted toward women. Second, the main risky behaviors that men engage in – violence and substance abuse – are not well captured in the data.
[25] Note: the tables for Mexico rural use 3 levels of rurality/urbanity: 1 = rural,, 0.5=from a small town, 0=urban. The information for a binary rural variable is reported above for comparison to Chile.

Chilean males in this age group and of this type live in rural areas. The "at-risk parents" are more rural than other groups (23% Chilean female cluster and 31% Mexican women are rural). The Chilean mothers that are early dropouts in this group report a high level of social exclusion, average family cohesion and are made up of 13% who self identify as indigenous. The "at-risk and inactive" clusters have the worst outcomes being idle (Tables 4e, 4f, 4i, 4j).

A small group of males – 7.5% of Chileans and 11.4% of Mexicans – categorized as risk type III come from relatively good backgrounds. The Chileans tend to be in the 18 to 24 age range, have caring families and good behaviors in general (best sexual behaviors of all the clusters) and yet are idle (Table 4i, column 2). None are married and few are registered to vote. These idle youth did not drop out of secondary school and are the youngest cluster in this cohort, being under 20 on average. Perhaps this reflects the difficulty some youth have in finding a job to settle into when first graduating– a closer look at the question that generated the work variable shows that 60% of these youth are looking for work. Overall, they are not very optimistic about the future, and do not feel very prepared for work. The Mexican men dropped out of school early, but most are working (99%). They were younger than most other clusters when they first started working and when they started their sexual activity and had children. Seventy-one percent are fathers and 99% are married (well above the other clusters), yet they are slightly more likely to have sexual relations with more than one person in the past year than they other clusters (Table 4e, column 3). Notably, this group is very small, suggesting that youth who have backgrounds with numerous protective factors can experience negative outcomes, but this is the exception; most youth who experience negative outcomes also have many risk factors in their lives.

Figures 1a and 1b show the average behavior and factor levels for each cluster type. First considering Type III, we see the clusters of this type have lower averages for those variables that might be considered protective factors and high values for those variables that might be considered risk factors.

No Risk

At the other end of the spectrum is 20-45 percent of the youth population that has neither engaged in risky behaviors, nor demonstrated a high level of potentially risky factors. There are fewer youth in this category in the older age groups than in the younger age groups (Tables 3a and 3b). In addition to the low presence of risk factors, these youth have a higher level of protective factors and good behaviors than those with type II or III risk levels (Figures 1a and 1b). All of the 12-14 and 15-17 years old youth with type 0 risk are in

school, few are working, none have children, and all initiated sexual behavior at a later age - indeed most have not initiated sexual behavior - than those in groups classified as risk types I, II, or III. Among the 18-24 year old groups, many with risk type 0 are still in school, fewer are working than in the other risk type groups, and fewer have started families. Generally speaking, these youth can be considered more economically advantaged than other youth, while youth in this group who are not from wealthy households have strong emotional connections with their families and feel connected to adults who care about them. They tend to have fewer risk factors, in terms of household abuse or exposure to drugs (Tables 4a-4j, last columns in each table).

Type I Risk

Those with a higher risk than type 0 but much better than risk type III are the 9 to 20 percent of the Chilean and Mexican samples that have many risk factors, but they are not engaged in risky behaviors nor suffering negative consequences, i.e. they can be classified as type I risk (Tables 3a and 3b). Their behaviors are similar to those with no risk despite having a higher level of risk factors. In some cases this is perhaps because of the level of protective factors in their lives, such as connectedness with adults outside of the household, extracurricular activities or spirituality, and in others resiliency is likely playing a part while others are on their way to graduating to type II risk. One group of Chilean males aged 18 to 24 (9.3% of the cohort) is resilient and yet has poor protective factors to make up for the poor family cohesion with some substance abuse and abuse in the home. This group did not drop out of school early, half are still in school, yet about 20% are idle and 10% report victimization (Table 4i, column 5). Rural youth are over represented in some clusters, and underrepresented in others. Across age groups, there are fewer youth in this category in the older age groups than in the younger age groups as seen in Tables 3a and 3b. These youth are either too young to have begun engaging in risky behaviors, or they may be considered resilient.

Some youth in this risk level (across all age groups, genders and countries) report a high level of social exclusion and can be considered loners. Those that report a high level of social exclusion and have good behaviors tend to report a higher level of connectedness with their families (better relationships, doing more activities with their parents, better cohesion, etc.), suggesting that being socially excluded from their peers could indeed be a protective factor in some cases. For others, though, the social exclusion may lead to deviant behaviors, as demonstrated by very high levels of social exclusion among older youth of risk type III.

The level of risk factors, protective factors and behaviors for this type of risk can be seen in Figures 1a and 1b for Mexico and Chile respectively. Behaviors are better than for types II and III risk and comparable to type 0 risk. Risk factors are worse than those with type 0 risk, but better than those with types II and III risk. Protective factors in Mexico are comparable to those with no risk; however those who are classified as type I risk in Chile have the fewest protective factors of any risk type, suggesting that this is a temporary phase with potentially worse to come.

Type II Risk

The most heterogeneous of the risk groups are the youth in clusters of risk type II who are 28% of the Chile sample and 25% of the Mexico sample. In Mexico, they are younger than the risk type III youth, with 33 percent of Mexicans age 12 to 14, 12 percent of Mexicans age 15 to 17, and 26.8 percent of Mexicans age 18 to 24 in this group. In contrast, about 20 percent of Chileans in the 15 to 17 age group and 33 percent in the 18 to 24 age group fit in this category. These youth are engaging in risky behaviors but are not suffering the consequences of their behaviors (Tables 3a and 3b). Figures 1a and 1b show that these youth have worse behaviors and outcomes compared to those with types 0 and I risk, however are not nearly as bad as those with type III.

The types of behaviors vary amongst the different age groups and genders. The 12 to 14 year olds have a high level of early workers yet are still in school, in contrast to the most at-risk groups which are not in school. The boys falling into this category started work on average much earlier than in the other clusters, including the type III youth. For these youth, early work does not appear to influence their other behaviors but since they are younger than type III youth, there is a chance that some will "graduate" to risk type III. On the other hand, the girls in this category did not start work as early as the most at-risk cluster, and not as many of them are working (Tables 4a and 4b, columns 2).

Among the 15-17 year olds, those with risk type II started sexual activity and work (more so for males) at a young age. They are still in school and are not parents, but their risky behaviors increase the likelihood that they will dropout or have children (Tables 4c, 4d, 4g, and 4h).

Among the 18 to 24 year olds, a wide range in behavior for youth with type II risk is observed. Some have high levels of many risky activities, while others strongly display only one risky behavior (social exclusion or risky sex). A special group that breaks out can be identified as "resilient" youth who display type II risk, but have the luck, genetics, or

psychological make-up such that they have not suffered the consequences of their actions. For example, the resilient groups might have started work at an early age and yet managed to graduate from high school (Tables 4e, column 5; Table 4f column 5; Table 4j column 2).

These youth tend to come from families with low parental education (though not as low as the most at-risk clusters), limited access to healthcare, and poor relationships with their parents. The 12 to 14 year olds have parents with low education and have poor relationships with their parents, but they have better relationships with their parents and a higher level of spirituality than those classified as risky type III. These youth may be on their way to the risk type III later in life or they may be resilient to the type I and II risk present in their lives. In the 15 to 17 age range, the different sub-groups have high levels of risk factors (low parental education, limited access to healthcare, and poor family cohesion) and medium levels of protective factors. The females tend to come from particularly difficult family backgrounds but have a high level of connectedness, often with an adult outside the family. The youth in the 18 to 24 age group tend to be more connected or resilient than those with type III risk, or have a significantly lower level of risk factors present (Tables 4a-4j).

B. Correlations among Risk Factors, Protective Factors, Risky Behaviors, and Outcomes

Plotting risk or protective factors against behaviors and estimating the correlation confirms that young people with worse personal, micro, and macro influences in their lives are more at-risk than those who have positive influences. To keep this discussion short, we will discuss these relationships only for the 18 to 24 year old cohort in each country. The general trends described here are also observed when analyzing the younger cohorts as well as when grouping the clusters by each variation of age/gender groups.[26]

Good behaviors are positively correlated with protective factors in both Mexico and Chile as seen in Figures 2a and 2b respectively. The relationship is significant in Mexico for all age cohorts individually as well as when grouped together, however is not significant in Chile. The weaker results in Chile are potentially due to the lower number of available protective factors in its dataset. The analysis was also performed on youth aged 12 to 29 in

[26] These results are available from the authors upon request. While the relationships are not significantly different from zero in all cases due to the limited number of data points once disaggregating by gender and age, the relationships between factors and behaviors are clearer. Similarly, when grouping all clusters together for all ages and both genders, the results are not always interesting due to the wide difference in behaviors amongst age groups.

both countries as a robustness check[27]. Results were significantly different than 0 (at a 5% level) with this wider age range in both countries, suggesting the importance of age in some behaviors and outcomes.

As suggested by the ecological framework, the opposite results are observed when considering the correlation between risk factors and positive behaviors (Figures 3a and 3b) for 18-24 year old youth. The relationships are significant in Mexico at a 10% level and not significant in Chile at the 10 percent level. The relationship is strengthened (1% level) in Mexico when grouping all clusters together rather than breaking up the different age groups. This is partially due to few risk factors in the Mexican data as compared to protective factors. The Chilean dataset, which is richer in risk factors than in protective, shows a negative and significant relationship at 1% between risk factors and positive behaviors in the 15 to 17 year old age group, but only at 15% in the 18 to 24 year olds[28].

C. **Proxies for At-Risk**

Due to the nature of the data used in this analysis, we can use a large set of risk and protective factors to predict the level of risky behavior that will be displayed by youth. Some of our variables are easily observable characteristics and are regularly found in most household data sets, but most of these data are unique and not available for most countries and youth. Thus, it would be useful to identify a youth's risk level by a single or small set of easily observable characteristics. We test a set of observable variables to determine their usefulness in predicting risky behavior: poverty, living in a rural area, ethnicity, gender, and age.

Poverty is a fairly good proxy to identify youth at-risk. Tables 3a and 3b summarize the information from the clusters and reveal that as the risk type rating of a group increases, the average level of parental education falls. Plotting the clusters against wealth proxies and estimating the slope of the line confirms this relationship.[29] Figure 4a shows that in Mexico, the more at-risk clusters have lower parental education, and the relationship is statistically significant at the 1% level. The same result emerges with each Mexico age group as well as when merging the Mexico age groups together, with a negative and significant coefficient on

[27] Results are available upon request.
[28] The relationship was significant at 10% and negative for the 18 to 29 year old cohort.
[29] Since poverty level is an exogenous variable, we test these relationships for all poverty indicators: education level of household head, household ownership of luxury or durable goods, or monthly earnings of heads of households. Since there are no significant differences between the various proxies for poverty, we chose parental education level since it is a widely available and observable indicator.

poverty for all age groups. The coefficients are larger negative values for the younger age groups. Similarly, Chile's graph of poverty against positive behaviors in Figure 4b shows a statistically significant downward trend of behavior against poverty (proxied by parental education).[30] This negative trend is significant for the younger cohort as well, with an even larger magnitude.

While poverty is a reasonable candidate as an indicator in identifying at-risk groups,[31] we cannot conclude that economic class is the sole source of clustering since a close look at the clusters shows that not everyone in the at-risk groups is poor. The figures in Annex C use the density of the various possible general levels of education of heads of households for each cluster and demonstrate that poverty is correlated with risk level, however each risk cluster has youth with parents of varying education levels. Furthermore, we do not suggest that poverty causes risky behavior, but, poverty, for whatever reason, can be used to identify ex ante which youth have a higher propensity for engaging in risky behaviors.

Living in a rural area is a candidate for a proxy for risk; in both countries, rural youth are over 10 percentage points more likely to be in risk types I, II, and III than are urban youth, and ten percentage points less likely to be risk-free (Tables 3a and 3b). Figures 5a and 5b confirm this correlation, demonstrating that living in a rural area is negatively correlated with positive behaviors and outcomes at the 1% level. That said it is important to note that not all rural youth are at risk, and that most at-risk youth live in urban areas, since more of the general population lives in urban areas.[32] Again, in Chile living in a rural area may be more risky than in Mexico, since the more at-risk clusters in Chile are more rural compared to the national average than Mexico.

Ordered logistic regressions relating the risk level to poverty level and living in a rural area (when controlling for age and gender as well as when not doing so) demonstrate a large, significant and negative relationship for both variables as shown in Table 5. This

[30] Similarly Chile had two variables available to represent poverty: parent/caregiver education level and the economic class variable from the survey data (nse – nivel socio-economico). The "nse" variable was constructed by the Chilean National Institute of Youth using parental/caregiver education and occupation or an average of six durable goods in the event that the other information was not available.
[31] Indeed, Blakely et al. (2005) used income level to estimate relative risk levels. Their risk factors are: tobacco, drinking alcohol, not having access to safe water/sanitation, indoor air pollution exposure and obesity.
[32] In Mexico, the rural indicator was constructed to equal 1 if rural (< 2500 inhabitants), 0.5 if living in a small to medium sized town (2500 to 99,000) and 0 if living in a large urban setting. The rural indicator in Chile equals 1 if rural with more than 2000 and less than 5000 inhabitants and 0 if urban with more than 5000 inhabitants.

suggests that using both poverty as well as rurality might prove useful in identifying the at-risk youth.

Ethnic identity is not a good proxy for youth at-risk, although indigenous youth are over-represented among clusters defined as risk type III (Table 3b). A closer look at the indigenous youth in the Chile sample shows that while indigenous youth have worse risk factors and behaviors, they have better protective factors.[33] The more at risk clusters are about 5 percentage points more indigenous than the average, but the negative relationship between risk factors and positive behaviors is slight and insignificant as seen in Figure 6.[34] We must remember, however, that this is a self-reported survey, with the youth indicating if they identify with a particular indigenous group.[35]

Overall levels of risky behavior increase with age, as seen in Tables 3a and 3b. The bubble graphs by age confirm this observed correlation.[36] Older youth have lower levels of positive behaviors across the board. On the other hand the poverty level for the most at risk clusters is higher for the younger cohorts than the older, indicating that with age a wider variety of background characteristics are involved in risky behaviors.

In Mexico, gender does not seem to matter (Table 3a) while in Chile women seem to be less at-risk (Table 3b). A closer look at the data reveals that male risky behavior is related to education and employment and females are more at risk for poor sexual behavior at younger ages. The females participating in risky behaviors at young ages tend to exhibit many risky behaviors, more than the males in the same cohort. Unfortunately, we were not able to look at violent or criminal behavior for this study, which is generally considered to be more common among men. In both Chile and Mexico, the most at-risk groups of females are more rural than males, despite the fact that similar percentages of males and females (across all age ranges) are rural. So being female and living in a rural area is more risky than being male and living in a rural area[37].

Finally, we see a larger percentage of youth at risk in Mexico than in Chile, perhaps reflecting overall macroeconomic growth or government policy.

---

[33] Further details available upon request.
[34] Youth 15 to 17 years of age have a similar plot which is available upon request.
[35] See Hall and Patrinos (2006) for documentation of the limitations of using self-reported ethnicity.
[36] This is seen with both the Chile and the Mexico clusters by comparing the different age cohorts. We do not present the graphs here but they are available upon request.
[37] Approximately 13% of females and males live in rural areas in Chile and 24.5% in Mexico. Of those with type III risk, 20% of males and 23.6% of females in Chile and 29.7% of males and 33.7% of females in Mexico live in rural areas.

VI. Conclusion

Based on unique youth surveys in Chile and Mexico, we find that more than half of young people in these two countries can be considered at-risk. In Mexico, nearly one-third of people age 12-24 are suffering the consequences of negative youth behaviors – adolescent mothers, early school dropout, not working – while 17 percent of Chilean youth are in this situation. They also come from the poorest families and have the fewest social bonds. Another one-quarter of Mexican and Chilean youth are engaging in negative behaviors and on their way to the worst-off category. A younger group in Mexico (20 percent) and Chile (8 percent) are not engaging in risky behaviors, but they have factor in their lives that suggest that they may be graduating to these more harmful groups before long since they also lack the social supports and mental health that the no-risk group boasts. These findings confirm the U.S. literature that negative factors in a young person's life is highly correlated with negative behaviors. However, further analysis, preferably with panel data, is needed to understand the transition more clearly.

Some factors repeatedly arise in the best-off or worst-off groups. Positive relationship with the family is a recurring protective factor across the clusters. The clusters show that youth who live(d) with both parents have a lower incidence of all risky behaviors. Those youth who feel connected to a parent, i.e. those who feel that they can relate to a parent, the parent cares for them, they can depend on the parent, etc, also have lower risky behavior and negative outcomes than youth who do not feel connected. Connection with non-family members can partially compensate for absent parental connection, as shown by youth who are connected to non-family members having less risky behavior than those who do not have connections with anyone. Conversely, family abuse, substance abuse and lack of family cohesion cluster with victimization in Chile, while non-positive feedback from parents is correlated with risky behaviors in Mexico.

Positive institutional factors are also correlated with positive behaviors. School quality matters in Chile, and an individual's relationship with his/her community (trust, feeling of school quality) is positively correlated with voter registration. Low spirituality/church attendance clusters loosely with the risk type III groups. Surprisingly, social exclusion is rarely correlated with negative behaviors. Instead, the "loner" youth usually showed up as sub-groups in the less at-risk clusters.

Negative behaviors (or outcomes) cluster together. An early age of starting work clusters loosely with an early age at onset of sexual activity for boys and girls. Other risky behaviors by females cluster with risky sexual behavior while the pattern is weaker for males:

For those 12 to 17 and for all females, inactivity and early school drop cluster, which is not the case for the older males. For women 18 to 24 years of age, inactivity and marriage cluster together with a high incidence of having a child, however in this age range this is not risky behavior.

There is no simple answer as to which variables can be used to proxy being at-risk, however poverty is the best option of those available. Poor youth are more at risk on average. Rural youth are more at risk on average than urban youth.

When considering prevention programs to target at risk youth, it is important to remember that risk profiles vary with youth development. Young youth not exhibiting any risky behavior are not out of danger. Therefore age appropriate programs targeting those with type I risk could potentially prevent risky behaviors later in life.

In addition, intervention plans should take gender into account. Gender appears to matter more so with regards to sexual health as younger at-risk females tend to participate in all risky behaviors available in our dataset, while young males exhibit some risky behaviors but not all. The public health literature suggests that perhaps violent behavior is more important for boys however this could not be tested with these data used in this study.

Table 1 – Construction of Variables

	Risk Type I, II, III	If variable appears in the data set		Used in the cluster analysis
		Chile	Mexico	
Behaviors/outcomes				
Not idle – either in school nor working	II, III	X	X	X
No early school dropout (i.e. completed secondary school)	II	X	X	X
Literate (can read and/or write a message)	III		X	X
In school	II, III	X	X	
Years of education completed	I, III	X	X	
Working (Not Working[c])	II, III	X	X	
Older age when started working	II, III	X	X	X
Safe sex (among the sexually active)	II	X	X	
Low number of sexual partners in the past year	II		X[a]	
Older age at onset of sexual activity	II	X	X[a]	X
Older age at first pregnancy/child	III	X	X[a]	X
Has at least one child (Does not have at least one child[c])	III	X	X	
Married (Not Married[c])	I, III	X	X	X
Participate in extracurricular activities	II	X	X	X
Registered to vote, planning on voting in the next election, desires to vote when of age (if under 18 years old)	III	X	X[b]	X
Has not been a victim of a crime (proxy for criminal activity)	II	X		X
Attitude towards drugs (respondent can justify using drugs)	II		X[a]	
Attitude towards alcohol (respondent can justify getting drunk)	II		X[a]	
Protective Factors				
Trust in governmental institutions	I	X[d]	X[a,e]	X
Trust in community institutions	I	X[f]	X[a,g]	X
Connected (whether youth reaches out – for talk or help- to someone when they have problems)	I	X	X	X
Living with both, one or no parents	I		X	X
Positive relationship with father	I	X[h]	X[i]	X
Positive relationship with mother	I	X[j]	X[k]	X
Connected with an adult other than parents	I	X		X
Communication with parents (talk to parents when facing a personal problem)	I		X	
Church attendance	I	X	X	X
Spiritual influence in beliefs, opinions and attitudes	I		X	
School quality[l]	I		X	X
Feeling optimistic about future work	I	X		
Feeling prepared for future employment	I	X		
Sense of wellbeing (level of happiness reported by the youth)	I		X[a]	
Risk Factors				
Have felt discriminated against	I	X		
Parental education – this is a good proxy for poverty	I	X	X	X
Household ownership of durable/luxury goods[m]	I		X	
Monthly earnings of heads of household	I		X	
Limited access to healthcare	I		X	X

Rural residence (versus urban)	I	X	X	
Social exclusion	I, III	X	X	X
Poor family cohesion	I	X		X
Abuse in the home	I	X		X
Substance abuse in the home	I	X		X
Level of perceived violence in the neighborhood	I	X		X
Indigenous (self-identifying with an indigenous group)	I	X		
Parental influence with regards to smoking and alcohol[n]	I		X	X
Parental response to misbehavior[o]	I		X	
Parental response to good behavior[p]	I		X	

[a] for respondents age 15 or older
[b] for respondents age 18 or older
[c] for respondents under age 18
[d] indicates level of confidence in government, congress, city government, political parties, judicial system
[e] indicates level of confidence in politicians, judges, the police, and the military
[f] level of confidence in hospitals, the Catholic Church, schools, universities, and family
[g] indicates level of confidence in teachers, doctors, shop owners, union leaders and priests
[h] quality of relationship with father on various attributes (communication, demonstration of love or affection, understanding and help with problems, respect for private life of youth, the time spend with father)
[i] variety of topics that the youth communicates with the father about (school, politics, religion, sexual relations, work, and other topics)
[j] indicates quality of relationship with mother on various attributes (communication, demonstration of love or affection, understanding and help with problems, respect for private life of youth, the time spend with mother)
[k] variety of topics that the youth communicates with the mother about (school, politics, religion, sexual relations, work, and other topics)
[l] rank of the overall quality of the youth's current/past school as reported by youth (physical building, scholastic materials, teachers preparation, content of courses and teachers assistance)
[m] includes radio recorder, CD burner, TV, cable, VCR, game console, telephone, computer, internet, car/truck/van - could be used instead of household education level to indicate economic class
[n] indicates level of control parents attempt to control children's behaviors (do they forbid smoking/drinking, grant periodic permission, allow the child to make his/her own decision)
[o] how parents respond when child bothers/angers them (0=by talking with their child, 0.5 = punishing, 1 = beating/hitting, insulting, accusation in front of others, stop talking)
[p] indicates frequency with which parents use positive feedback (words of encouragement, hug/kiss, give a gift, concede to something) when child does something good/correct (0 = always, 0.5 = sometimes, 1 = never)

Table 2 – Descriptive Statistics

	Mexico				Chile		
Actual percentages	All ages	12 to 14	15 to 17	18 to 24	All ages	15 to 17	18 to 24
Average age	17.19	13.0	16.0	20.8	18.9	15.95	20.8
Percent female	53.5	50.65	52.9	55.9	53.0	50.5	54.6
Percent indigenous	-	-	-	-	10.8	11.5	10.4
Percent rural	24.7	28.1	26.0	21.6	13.2	13.3	13.1
Behaviors and Outcomes							
Share not completing high school	29.9	10.0	33.0	42.3	14.2	8.4	17.8
Share in school	45.3	85.9	61.8	27.5	60.5	91.4	41.0
Share idle/inactive	20.3	8.9	17.7	29.9	27.3	6.7	35.9
Share working	41.9	17.2	35.1	35.1	21.4	4.7	32.0
Share having sex	52.1	-	13.3	56.0	59.1	26.9	79.4
Share of the sexually active using protection	50.6	-	54.0	52.5	61.9	59.9	62.3
Share reporting at least 1 child	19.4	-	2.5	28.9	18.6	3.9	27.9
Share married	15.7	0.4	4.1	32.9	4.8	0.1	7.7
Risk and Protective Factors							
Share reporting physical or psychological abuse in home	-	-	-	-	6.3	6.2	6.4
Share reporting problems arising from substance abuse in home	-	-	-	-	8.5	7.0	9.5
Share without access to medical services	49.4	50.2	50.7	48.0	-	-	-
Share attending church weekly	9.3	11.3	10.3	7.3	21.6	20.9	18.1
Share attending church at least once in the past month	66.3	72.0	65.9	62.5	-	-	-
Share believing in God	-	-	-	-	94.8	95.1	94.6
Share optimistic about the future	-	-	-	-	87.7	88.2	87.3
Share reporting being happy (sense of well-being)	98.2	-	98.0	98.4	-	-	-
Share reporting social exclusion	53.2	66.8	53.9	43.1	17.7	10.1	22.5
Share with parents who have a primary degree or less	56.1	56.3	58.0	55.0	38.1	38.8	37.7
Share reporting they have felt discriminated against	-	-	-	-	54.7	54.2	55.1

Table 3a – Mexico Cluster analysis overview

	Type III	Type II	Type I	None
Percent of total sample	33.1	25.1	20.2	21.6
Percent of 12-14 year olds	14.1	33.1	24.3	28.5
Percent of 15-17 year olds	38.9	11.9	27.5	21.7
Percent of 18-24 year olds	43.4	26.8	13.1	16.7
Percent of males	31.9	23.2	22.2	22.7
Percent of females	34.2	26.7	18.4	20.7
Average Parental Education (poverty proxy)	Primary graduate, some secondary	Primary graduate, some secondary	Secondary graduate	Secondary graduate
Percent of rural	42.9	24.4	18.1	14.6
Percent of urban	29.9	25.3	20.8	23.9

Table 3b – Chile Cluster analysis overview

	Type III	Type II	Type I	None
Percent of total sample	16.8	28.0	8.7	46.5
Percent of 15-17 year olds	8.3	20.4	15.9	55.5
Percent of 18-24 year olds	22.2	32.8	4.2	40.8
Percent of males	23.7	19.1	13.2	44.1
Percent of females	10.8	4.7	35.9	10.8
Average Parental Education (poverty proxy)	Primary Graduate	Some Secondary Education	Some Secondary Education	Secondary Graduate
Percent of rural	27.2	29.5	7.0	36.3
Percent of urban	15.3	27.7	9.0	48.0
Percent of indigenous	21.7	26.3	8.7	43.3
Percent of non-indigenous	16.3	28.2	8.7	46.9

Table 4a: Cluster analysis results: Mexican Males Aged 12 to 14

Cluster name	**at-risk**	**very early workers**	**advantaged loners**	**advantaged youth**
age	13.25	13.07	12.87	13.05
Behaviors Average	**0.41**	**0.69**	**0.83**	**0.77**
older age when first started working	0.42	0.11	1.00	0.62
not idle	0.49	1.00	1.00	1.00
no early school dropout	0.01	1.00	1.00	1.00
literacy	0.91	1.00	1.00	1.00
not married	0.99	1.00	1.00	1.00
participate in activities	0.14	0.29	0.17	0.23
not working	0.51	0.58	0.99	0.81
in school	0.01	1.00	1.00	1.00
years of education completed	0.20	0.28	0.27	0.28
Protective Factors Average	**0.54**	**0.58**	**0.60**	**0.58**
relationship with father	0.37	0.42	0.40	0.40
relationship with mother	0.42	0.48	0.46	0.46
connected	0.96	0.84	1.00	1.00
live with both parents	0.87	0.90	0.93	0.90
church attendance	0.36	0.41	0.39	0.39
school quality	0.75	0.81	0.83	0.82
communication with parents	0.49	0.52	0.55	0.51
spiritual influence	0.15	0.22	0.15	0.17
Risk Factors Average	**0.57**	**0.57**	**0.54**	**0.41**
parental influence (alcohol & smoking)	0.44	0.38	0.35	0.37
social exclusion	0.58	1.00	1.00	0.00
limited access to healthcare	0.63	0.52	0.45	0.47
parental response to misbehavior	0.38	0.31	0.29	0.32
parental response to good behavior	0.58	0.52	0.51	0.53
rural	0.55	0.49	0.45	0.43
parental education	0.80	0.70	0.67	0.67
household ownership of goods	0.84	0.79	0.79	0.77
monthly earnings household heads	0.85	0.80	0.78	0.78
n	758	1193	1893	1989
Risk Type	III	II	I	-
Percent of sample	13.0%	20.5%	32.5%	34.1%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4b: Cluster analysis results: Mexican Females Aged 12 to 14

Cluster name	**at-risk**	**early workers**	**advantaged loners**	**advantaged youth**
Age	13.33	12.98	12.97	13.13
Behaviors Average	**0.43**	**0.79**	**0.83**	**0.80**
older age when first started working	0.66	0.74	1.00	0.77
not idle	0.28	1.00	1.00	1.00
No early school dropout	0.01	1.00	1.00	1.00
literacy	0.92	1.00	1.00	1.00
not married	0.97	1.00	0.99	1.00
participate in activities	0.13	0.24	0.24	0.25
not working	0.73	0.87	0.99	0.89
in school	0.01	0.99	1.00	1.00
years of education completed	0.20	0.27	0.29	0.29
Protective Factors Average	**0.56**	**0.59**	**0.61**	**0.59**
relationship with father	0.36	0.39	0.42	0.40
relationship with mother	0.45	0.50	0.52	0.51
Connected	0.95	0.93	1.00	1.00
live with both parents	0.86	0.88	0.93	0.90
church attendance	0.45	0.44	0.42	0.45
school quality	0.77	0.82	0.81	0.81
communication with parents	0.45	0.50	0.55	0.48
spiritual influence	0.18	0.17	0.17	0.18
Risk Factors Average	**0.66**	**0.65**	**0.48**	**0.48**
parental influence (alcohol & smoking)	0.40	0.48	0.06	**0.34**
social exclusion	0.73	1.00	**1.00**	0.00
limited access to healthcare	0.64	0.69	0.00	**0.43**
parental response to misbehavior	0.30	0.27	0.26	0.27
parental response to good behavior	0.59	0.51	0.47	0.51
Rural	0.59	0.49	0.35	0.40
parental education	0.81	0.71	0.60	0.65
household ownership of goods	0.86	0.81	0.76	0.76
monthly earnings household heads	0.87	0.81	0.74	0.75
N	908	2723	982	1376
Risk Type	III	II	I	-
Percent of sample	15.2%	45.5%	16.4%	23.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4c: Cluster analysis results: Mexican Males Aged 15 to 17

Cluster name	**at-risk**	**becoming at risk**	**loners**	**advantaged youth**
age	16.16	15.89	15.88	15.97
Behaviors Average	**0.64**	**0.84**	**0.85**	**0.84**
older age when first started working	0.25	0.41	0.52	0.48
not idle	0.74	1.00	1.00	1.00
no early school dropout	0.06	1.00	1.00	1.00
literacy	0.97	1.00	1.00	1.00
older age at onset of sexual activity	0.85	0.92	0.90	0.88
older age at first pregnancy	0.98	1.00	1.00	1.00
not married	0.97	1.00	1.00	1.00
participate in activities	0.17	0.27	0.27	0.32
not working	0.30	0.63	0.73	0.73
in school	0.06	0.99	0.99	1.00
years of education completed	0.28	0.39	0.42	0.42
safe sex	0.85	0.93	0.91	0.90
low number of sexual partners in past year	0.96	0.98	0.98	0.97
does not have a child	0.99	1.00	1.00	1.00
attitude towards drugs	0.97	0.97	0.96	0.95
attitude towards alcohol	0.88	0.89	0.91	0.85
Protective Factors Average	**0.54**	**0.58**	**0.59**	**0.58**
Connected	0.91	1.00	1.00	1.00
live with both parents	0.84	0.88	0.89	0.88
relationship with father	0.38	0.42	0.44	0.43
relationship with mother	0.44	0.47	0.50	0.49
trust in governmental institutions	0.35	0.36	0.36	0.38
trust in community institutions	0.58	0.60	0.60	0.61
church attendance	0.32	0.35	0.35	0.32
school quality	0.78	0.81	0.81	0.82
communication with parents	0.44	0.48	0.51	0.47
spiritual influence	0.19	0.19	0.18	0.18
sense of wellbeing	0.82	0.85	0.89	0.87
Risk Factors Average	**0.54**	**0.57**	**0.45**	**0.29**
parental influence (alcohol & smoking)	0.54	0.46	0.42	0.45
social exclusion	0.46	0.48	1.00	0.00
limited access to healthcare	0.61	1.00	0.04	0.00
parental response to misbehavior	0.29	0.27	0.21	0.24
parental response to good behavior	0.60	0.55	0.50	0.53
rural	0.51	0.49	0.35	0.29
parental education	0.78	0.72	0.59	0.56
household ownership of goods	0.82	0.78	0.73	0.69
monthly earnings household heads	0.85	0.82	0.74	0.70
n	1765	1120	681	869
Risk Type	III	II	I	-
Percent	39.8%	25.3%	15.4%	19.6%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4d: Cluster analysis results: Mexican Females Aged 15 to 17

Cluster names	early workers"	at-risk mothers & wives	at-risk idle dropouts	resilient loners	advantaged loners	advantaged youth
age	16.20	16.42	16.07	15.85	15.85	15.92
Behaviors Average	**0.66**	**0.46**	**0.68**	**0.86**	**0.88**	**0.88**
older age when started working	0.26	0.43	0.65	0.62	0.71	0.67
not idle	1.00	0.21	0.02	1.00	1.00	1.00
no early school dropout	0.00	0.07	0.00	0.96	0.99	1.00
literacy	1.00	0.96	0.95	1.00	1.00	1.00
older age at onset of sexual activity	0.97	0.29	0.97	0.99	0.98	0.97
older age at first pregnancy	1.00	0.43	0.99	1.00	1.00	1.00
not married	0.99	0.34	0.99	1.00	0.99	1.00
participate in activities	0.16	0.07	0.15	0.25	0.27	0.31
not working	0.00	0.83	0.98	0.77	0.86	0.83
in school	0.00	0.06	0.00	0.95	0.99	1.00
years of education completed	0.27	0.27	0.26	0.38	0.42	0.42
safe sex	0.96	0.19	0.96	0.99	0.98	0.97
low number of sexual partners in past year	0.99	0.89	0.99	1.00	1.00	0.99
does not have a child	1.00	0.46	0.99	1.00	1.00	1.00
attitude towards drugs	0.98	0.97	0.97	0.97	0.97	0.96
attitude towards alcohol	0.94	0.94	0.94	0.95	0.93	0.89
Protective Factors Average	**0.57**	**0.49**	**0.57**	**0.57**	**0.62**	**0.60**
connected	1.00	0.85	0.96	0.82	1.00	1.00
live with both parents	0.85	0.27	0.87	0.85	0.91	0.86
relationship with father	0.43	0.41	0.45	0.48	0.52	0.45
relationship with mother	0.39	0.35	0.36	0.41	0.44	0.42
trust in governmental institutions	0.35	0.32	0.34	0.33	0.36	0.37
trust in community institutions	0.59	0.57	0.58	0.58	0.60	0.62
church attendance	0.20	0.16	0.18	0.22	0.20	0.21
school quality	0.82	0.79	0.78	0.82	0.82	0.80
communication with parents	0.51	0.46	0.46	0.53	0.57	0.54
spiritual influence	0.42	0.31	0.45	0.45	0.43	0.42
sense of wellbeing	0.81	0.85	0.79	0.82	0.89	0.86
Risk Factors Average	**0.54**	**0.53**	**0.57**	**0.62**	**0.43**	**0.35**
parental influence (alcohol & smoking)	0.43	0.43	0.41	0.40	0.38	0.39
social exclusion	0.60	0.39	0.68	1.00	1.00	0.00
limited access to healthcare	0.62	0.63	0.64	0.92	0.01	0.39
parental response to misbehavior	0.26	0.38	0.28	0.25	0.22	0.24
parental response to good behavior	0.59	0.58	0.58	0.52	0.47	0.49
rural	0.48	0.50	0.61	0.50	0.33	0.32
parental education	0.78	0.77	0.80	0.72	0.60	0.60
household ownership of goods	0.83	0.84	0.84	0.80	0.74	0.72
monthly earnings household heads	0.85	0.85	0.86	0.83	0.74	0.73
n	603	363	932	957	947	1178
Risk Type	III	III	III	I	I	-
Percent	12.1%	7.3%	18.7%	19.2%	19.0%	23.7%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4e: Cluster analysis results: Mexican Males Aged 18 to 24

Cluster name	**idle**	**drinkers**	**working dads**	**loners**	**resilient loners**	**advantaged students**
age	20.58	19.60	22.28	21.19	20.96	20.15
Behaviors Average	**0.47**	**0.56**	**0.64**	**0.58**	**0.63**	**0.69**
older age when started working	0.42	0.30	0.30	0.32	0.40	0.46
not idle	0.00	0.98	1.00	1.00	1.00	1.00
no early school dropout	0.27	0.51	0.11	0.00	0.25	1.00
literacy	0.99	0.90	1.00	1.00	1.00	1.00
older age at onset of sexual activity	0.66	0.69	0.50	0.62	0.71	0.68
older age at first pregnancy/parenthood	0.91	0.92	0.69	0.92	0.95	0.98
married	0.14	0.14	0.99	0.05	0.06	0.01
participate in activities	0.19	0.18	0.18	0.20	0.24	0.31
registered to vote	0.95	0.95	0.98	0.96	0.96	0.93
working	0.00	0.81	0.99	0.91	0.78	0.56
in school	0.00	0.30	0.07	0.20	0.38	0.74
years of education completed	0.36	0.34	0.37	0.35	0.44	0.69
safe sex	0.57	0.61	0.22	0.51	0.63	0.64
low number of sexual partners in past year	0.89	0.90	0.87	0.87	0.90	0.89
has a child	0.16	0.14	0.71	0.15	0.11	0.03
attitude towards drugs	0.82	0.84	0.86	0.82	0.85	0.77
attitude towards alcohol	0.73	0.06	1.00	1.00	1.00	1.00
Protective Factors Average	**0.54**	**0.53**	**0.53**	**0.55**	**0.55**	**0.59**
relationship with father	0.46	0.45	0.48	0.47	0.49	0.50
relationship with mother	0.40	0.40	0.43	0.42	0.43	0.48
connected	0.96	0.95	0.99	1.00	0.81	1.00
live with both parents	0.75	0.73	0.34	0.73	0.79	0.83
trust in governmental institutions	0.34	0.36	0.35	0.34	0.38	0.36
trust in community institutions	0.59	0.60	0.62	0.61	0.63	0.64
church attendance	0.30	0.28	0.31	0.30	0.31	0.31
school quality	0.79	0.76	0.78	0.80	0.80	0.81
communication with parents	0.47	0.45	0.48	0.48	0.49	0.53
spiritual influence	0.19	0.20	0.19	0.17	0.20	0.19
sense of wellbeing	**0.82**	**0.83**	0.91	0.85	0.84	0.89
Risk Factors Average	**0.58**	**0.56**	**0.50**	**0.51**	**0.60**	**0.45**
parental influence (alcohol & smoking)	0.66	0.66	0.68	0.74	0.73	0.72
social exclusion	0.37	0.36	0.10	0.00	1.00	0.16
limited access to healthcare	0.61	0.52	0.38	0.46	0.40	0.44
parental response to misbehavior	0.28	0.29	0.31	0.29	0.24	0.22
parental response to good behavior	0.59	0.60	0.59	0.60	0.56	0.54
rural	0.48	0.44	0.38	0.36	0.38	0.24
parental education	0.74	0.73	0.66	0.73	0.69	0.54
household ownership of goods	0.78	0.80	0.80	0.76	0.75	0.66
monthly earnings household heads	0.84	0.83	0.80	0.82	0.81	0.71
n	743	1511	837	1778	1335	1142
Risk Type	III	III	III	II	I	None
Percent	10.1%	20.6%	11.4%	24.2%	18.2%	15.5%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4f: Cluster analysis results: Mexican Females Aged 18 to 24

Cluster name	at risk, idle married mothers	at risk, idle, and excluded married mothers	poor working early dropouts	somewhat advantaged married moms	resilient poor	advantaged students
age	21.35	21.09	21.27	21.48	19.02	20.27
Behaviors Average	**0.50**	**0.51**	**0.63**	**0.63**	**0.62**	**0.72**
older age when started working	0.53	0.56	0.47	0.46	0.48	0.63
not idle	0.00	0.07	1.00	1.00	0.95	0.89
no early school dropout	0.07	0.19	0.00	0.10	0.62	1.00
literacy	0.99	0.92	1.00	0.98	1.00	1.00
older age at onset of sexual activity	0.58	0.67	0.85	0.70	0.92	0.91
older age at first pregnancy/parenthood	0.60	0.68	0.87	0.76	0.97	0.96
married	0.63	0.40	0.12	0.45	0.03	0.05
participate in activities	0.11	0.11	0.18	0.18	0.22	0.29
registered to vote	0.98	0.98	0.97	0.98	0.97	0.97
working	0.00	0.05	0.83	0.82	0.60	0.40
in school	0.00	0.02	0.29	0.26	0.48	0.69
years of education completed	0.29	0.29	0.35	0.39	0.40	0.71
safe sex	0.36	0.50	0.75	0.54	0.88	0.86
low number of sexual partners in past year	0.90	0.93	0.96	0.93	0.98	0.98
has a child	0.68	0.56	0.23	0.40	0.05	0.06
attitude towards drugs	0.96	0.95	0.93	0.90	0.92	0.88
attitude towards alcohol	0.76	0.73	1.00	0.93	0.00	0.99
Protective Factors Average	**0.53**	**0.51**	**0.57**	**0.56**	**0.58**	**0.61**
relationship with father	0.43	0.43	0.46	0.45	0.46	0.51
relationship with mother	0.37	0.36	0.41	0.41	0.41	0.48
connected	1.00	0.64	0.85	1.00	1.00	0.95
live with both parents	0.26	0.52	0.74	**0.47**	0.80	0.80
trust in governmental institutions	0.35	0.35	0.35	0.35	0.34	0.36
trust in community institutions	0.61	0.61	0.62	0.62	0.61	0.63
church attendance	0.37	0.39	0.39	0.36	0.41	0.39
school quality	0.80	0.76	0.81	0.81	0.80	0.82
communication with parents	0.49	0.48	0.54	0.55	0.54	0.60
spiritual influence	0.18	0.20	0.21	0.22	0.22	0.25
sense of wellbeing	0.89	0.82	0.82	0.90	0.82	0.89
Risk Factors Average	**0.51**	**0.65**	**0.59**	**0.42**	**0.54**	**0.46**
parental influence (alcohol & smoking)	0.47	0.49	0.55	0.53	0.50	0.59
social exclusion	0.00	0.97	0.80	0.01	0.52	0.42
limited access to healthcare	0.51	0.65	0.57	0.14	0.54	0.36
parental response to misbehavior	0.36	0.33	0.28	0.30	0.26	0.21
parental response to good behavior	0.58	0.59	0.55	0.55	0.54	0.49
rural	0.45	0.54	0.40	0.26	0.40	0.24
parental education	0.71	0.75	0.74	0.63	0.72	0.55
household ownership of goods	0.82	0.83	0.79	0.76	0.78	0.68
monthly earnings household heads	0.84	0.86	0.84	0.75	0.82	0.72
n	1810	2324	1693	999	850	1644
Risk Type	III	III	II	II	I	-
Percent	19.4%	24.9%	18.2%	10.7%	9.1%	17.6%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4g: Cluster analysis results: Chilean Males Aged 15 to 17

Cluster name	**at-risk**	**risky sex**	**doing ok**	**loners**	**advantaged youth - some sexual activity**	**advantaged youth**
age	16.31	16.01	15.90	15.95	15.78	15.72
Behaviors/Consequences Average	**0.53**	**0.77**	**0.82**	**0.78**	**0.84**	**0.86**
older age when first working	0.39	0.52	0.67	0.71	0.74	0.97
not idle	0.36	1.00	1.00	1.00	1.00	1.00
no early school dropout	0.00	1.00	1.00	0.98	1.00	1.00
older age at onset of sexual activity	0.60	0.63	0.83	0.73	0.86	0.98
older age at first pregnancy/child	0.97	0.99	1.00	0.98	1.00	1.00
not married	1.00	1.00	1.00	1.00	1.00	1.00
Participate in activities	0.09	0.16	0.18	0.11	0.23	0.14
not a victim	0.92	0.91	0.91	0.91	0.93	0.94
not working	0.64	0.92	0.95	0.92	0.98	0.99
in school	0.00	1.00	1.00	0.98	0.99	1.00
years of education completed	0.31	0.46	0.45	0.46	0.45	0.45
safe sex	0.58	0.62	0.82	0.72	0.87	0.98
does not have a child	0.96	0.99	1.00	0.97	1.00	0.99
desires to vote	0.53	0.55	0.63	0.44	0.64	0.59
Protective Factors Average	**0.59**	**0.62**	**0.75**	**0.50**	**0.76**	**0.66**
relationship with father	0.65	0.62	0.70	0.70	0.92	0.86
relationship with mother	0.85	0.88	0.79	0.86	0.95	0.93
Connected to other adults	0.05	0.02	0.93	0.00	0.00	0.02
Connected	0.72	0.96	1.00	0.00	1.00	1.00
trust in government	0.31	0.33	0.37	0.28	0.44	0.37
trust in community	0.74	0.79	0.83	0.74	0.84	0.81
Spirituality	0.26	0.21	0.34	0.23	0.97	0.22
Risk Factors Average	**0.35**	**0.31**	**0.27**	**0.26**	**0.25**	**0.24**
poor family cohesion	0.38	0.38	0.28	0.35	0.19	0.26
abuse in the home	0.09	0.09	0.00	0.01	0.00	0.01
substance abuse in the home	0.08	0.19	0.00	0.00	0.00	0.00
social exclusion	0.13	0.01	0.00	0.11	0.00	0.17
community violence	0.15	0.21	0.25	0.18	0.29	0.19
felt discriminated against	0.61	0.63	0.51	0.50	0.46	0.48
live in a rural area	0.27	0.13	0.19	0.08	0.13	0.11
indigenous	0.20	0.12	0.13	0.10	0.17	0.08
low economic class (nse)	0.85	0.71	0.74	0.71	0.67	0.63
low parental education	0.75	0.59	0.61	0.59	0.56	0.50
N	75	359	88	105	126	265
Risk Level	III	II	I	I	-	-
Percent of sample	7.4%	35.3%	8.6%	10.3%	12.4%	26.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4h: Cluster analysis results: Chilean Females Aged 15 to 17

Cluster name	at-risk	high risk but connected	loners	advantaged youth	connected
age	16.26	16.15	15.91	15.94	16.00
Behaviors/Consequences Average	**0.44**	**0.79**	**0.81**	**0.83**	**0.82**
older age when started working	0.59	0.78	0.84	0.86	0.88
not idle	0.07	1.00	1.00	1.00	0.99
no early school dropout	0.01	1.00	1.00	1.00	1.00
older age at onset of sexual activity	0.37	0.67	0.84	0.89	0.87
older age at first parenthood	0.61	0.99	0.96	0.99	0.99
not married	0.98	1.00	1.00	1.00	1.00
participate in activities	0.02	0.12	0.08	0.11	0.11
not a victim	0.92	0.91	0.94	0.93	0.92
not working	0.93	0.93	0.99	0.99	0.98
in school	0.00	1.00	1.00	1.00	0.99
years of education completed	0.31	0.45	0.47	0.46	0.47
safe sex	0.28	0.63	0.82	0.88	0.84
does not have a child	0.51	0.98	0.95	0.98	0.99
desires to vote	0.62	0.55	0.50	0.55	0.49
Protective Factors Average	**0.61**	**0.60**	**0.55**	**0.66**	**0.74**
relationship with father	0.59	0.55	0.57	0.75	0.71
relationship with mother	0.75	0.71	0.85	0.91	0.86
connected to other adults	0.12	0.20	0.02	0.00	1.00
connected	0.84	1.00	0.41	1.00	1.00
trust in government	0.38	0.24	0.30	0.31	0.29
trust in community	0.79	0.70	0.70	0.81	0.77
spirituality	0.28	0.36	0.32	0.42	0.42
Risk Factors Average	**0.39**	**0.48**	**0.34**	**0.27**	**0.29**
poor family cohesion	0.39	0.69	0.39	0.33	0.38
abuse in the home	0.14	1.00	0.05	0.00	0.00
substance abuse in the home	0.13	0.20	0.05	0.05	0.06
social exclusion	0.55	0.10	0.53	0.00	0.07
community violence	0.25	0.29	0.34	0.31	0.33
felt discriminated against	0.52	0.82	0.55	0.51	0.58
live in a rural area	0.25	0.18	0.08	0.12	0.10
indigenous	0.11	0.20	0.12	0.10	0.08
low economic class (nse)	0.85	0.73	0.73	0.69	0.71
low parental education	0.72	0.59	0.59	0.56	0.58
n	95	60	133	661	90
Risk Level	III	II	I	-	-
Percent of sample	9.1%	5.8%	12.8%	63.6%	8.7%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4i: Cluster analysis results: Chilean Males Aged 18 to 24

Cluster name	idle dropout (dads)	idle high school graduates	working (fathers) - socially excluded	connected	not well protected	advantaged	advantaged students
age	20.99	19.98	21.02	20.65	20.43	20.41	21.72
Behaviors/Consequences Average	**0.37**	**0.37**	**0.48**	**0.48**	**0.48**	**0.51**	**0.56**
older age when started working	0.38	0.60	0.48	0.49	0.53	0.57	0.56
not idle	0.62	0.00	0.80	0.79	0.79	0.93	0.81
no early school dropout	0.28	1.00	0.80	0.85	0.99	0.98	0.93
older age at onset of sexual activity	0.41	0.54	0.53	0.48	0.47	0.51	0.48
older age at first parenthood	0.90	0.97	0.90	0.96	0.96	0.95	0.98
married	0.00	0.00	0.08	0.04	0.02	0.05	0.00
participate in activities	0.08	0.11	0.07	0.16	0.14	0.12	0.12
registered to vote	0.02	0.01	0.14	0.17	0.01	0.01	1.00
not victim	0.93	0.92	0.92	0.90	0.90	0.91	0.91
working	0.53	0.00	0.57	0.42	0.38	0.39	0.38
in school	0.13	0.00	0.28	0.44	0.52	0.64	0.56
years of education completed	0.43	0.61	0.55	0.61	0.62	0.65	0.68
safe sex	0.22	0.37	0.35	0.33	0.30	0.35	0.32
has a child	0.22	0.06	0.25	0.12	0.09	0.13	0.06
Protective Factors Average	**0.58**	**0.66**	**0.62**	**0.78**	**0.50**	**0.63**	**0.66**
relationship with father	0.56	0.85	0.69	0.74	0.61	0.68	0.72
relationship with mother	0.81	0.86	0.85	0.87	0.79	0.87	0.87
connected to other adults	0.02	0.00	0.02	1.00	0.00	0.03	0.00
connected	0.75	1.00	0.81	1.00	0.00	1.00	0.85
trust in government	0.27	0.34	0.29	0.33	0.28	0.29	0.39
trust in community	0.72	0.81	0.73	0.75	0.72	0.74	0.76
spirituality	0.19	0.32	0.28	0.39	0.21	0.27	0.35
Risk Factors Average	**0.36**	**0.25**	**0.38**	**0.27**	**0.28**	**0.25**	**0.25**
poor family cohesion	0.42	0.25	0.31	0.31	0.41	0.31	0.29
abuse in the home	0.05	0.00	0.02	0.00	0.09	0.08	0.01
substance abuse in the home	0.42	0.00	0.04	0.04	0.15	0.03	0.08
social exclusion	0.09	0.00	1.00	0.13	0.01	0.01	0.00
community violence	0.18	0.18	0.23	0.19	0.20	0.20	0.18
felt discriminated against	0.55	0.47	0.56	0.56	0.55	0.54	0.59
live in a rural area	0.22	0.16	0.17	0.09	0.09	0.10	0.11
indigenous	0.14	0.10	0.17	0.12	0.09	0.09	0.16
low economic class (nse)	0.83	0.75	0.74	0.68	0.69	0.65	0.59
low parental education	0.71	0.62	0.58	0.59	0.55	0.53	0.47
n	237	111	169	117	137	576	134
Risk Level	III	III	III	II	I	-	-
Percent of sample	16.0%	7.5%	11.4%	7.9%	9.3%	38.9%	9.0%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 4j: Cluster analysis results: Chilean Females Aged 18 to 24

Cluster name	early drop out mothers	resilient	spiritual	connected	advantaged students
age	21.74	20.80	20.83	20.86	20.49
Behaviors/Consequences Average	**0.31**	**0.44**	**0.45**	**0.48**	**0.56**
older age when first started working	0.54	0.57	0.60	0.60	0.71
not idle	0.12	0.55	0.36	0.63	0.93
no early school dropout	0.00	0.73	0.92	0.88	1.00
older age at onset of sexual activity	0.38	0.52	0.58	0.58	0.69
older age at first parenthood	0.50	0.75	0.78	0.80	0.93
married	0.30	0.05	0.18	0.09	0.01
participate in activities	0.02	0.04	0.04	0.07	0.07
registered to vote	0.05	0.05	0.05	0.09	0.16
not victim	0.94	0.91	0.93	0.93	0.93
working	0.12	0.24	0.24	0.33	0.29
in school	0.00	0.38	0.16	0.36	0.75
years of education completed	0.36	0.56	0.60	0.61	0.68
safe sex	0.12	0.30	0.34	0.39	0.53
has a child	0.88	0.46	0.45	0.39	0.15
Protective Factors Average	**0.58**	**0.59**	**0.61**	**0.72**	**0.64**
relationship with father	0.57	0.53	0.64	0.61	0.72
relationship with mother	0.72	0.73	0.84	0.82	0.86
connected to other adults	0.02	0.13	0.00	1.00	0.01
connected	0.86	0.89	0.79	1.00	1.00
trust in government	0.33	0.24	0.27	0.30	0.30
trust in community	0.75	0.73	0.75	0.74	0.76
spirituality	0.30	0.32	**0.37**	0.34	0.33
Risk Factors Average	**0.36**	**0.48**	**0.31**	**0.30**	**0.25**
poor family cohesion	0.32	0.59	0.31	0.30	0.30
abuse in the home	0.01	0.56	0.02	0.01	0.00
substance abuse in the home	0.00	0.67	0.01	0.01	0.00
social exclusion	0.58	0.33	0.46	0.34	0.02
community violence	0.32	0.29	0.27	0.29	0.29
felt discriminated against	0.52	0.70	0.55	0.58	0.51
live in a rural area	0.23	0.16	0.15	0.11	0.08
indigenous	0.13	0.10	0.08	0.09	0.10
low economic class (nse)	0.80	0.75	0.71	0.70	0.65
low parental education	0.66	0.62	0.58	0.56	0.52
n	209	210	582	160	622
Risk Level	III	II	II	II	-
Percent of sample	11.7%	11.8%	32.6%	9.0%	34.9%

Notes: The highlighted rows indicate that the variables were not included in the clustering exercise. They are exogenous veriables.

Table 5 – Ordered logistic regression for Risk Level

Mexico	High Risk Level	High Risk Level
Low parent education (1)	0.48	0.485
	(0.012)**	(0.012)**
Being rural (rural =1, urban=0)	0.401	0.303
	(0.023)**	(0.023)**
Age		0.135
		(0.003)**
Gender (female=2, male=1)		0.128
		(0.019)**
Observations	37903	37903

Standard errors in parentheses
* significant at 5%; ** significant at 1%

Figures

Figure 1a. Cluster Group Means – Mexico

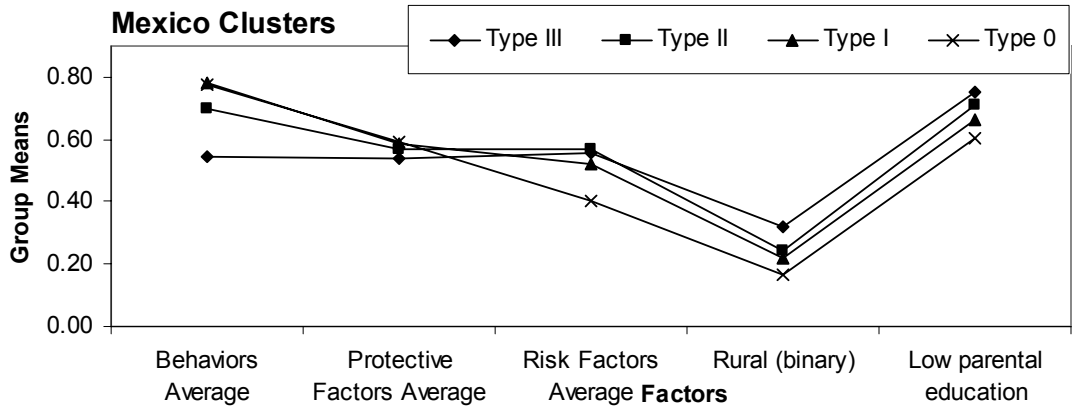
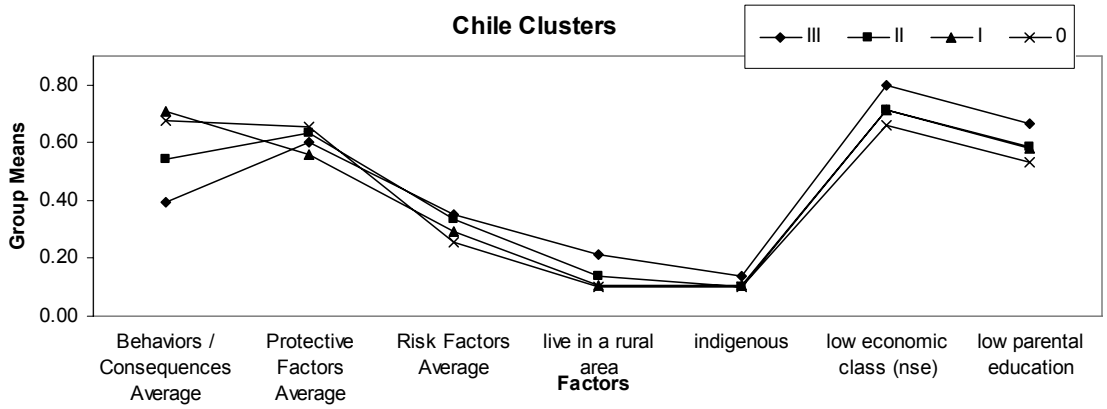


Figure 1b. Cluster Group Means – Chile



Note: The values of the group means are on a 0 to 1 scale as described in the variable construction section. Thus, the relative scale of the means between the different x values (primarily factors) is not interpretable. What is important is the relationship between the graphs for each risk level. As seen above, the Type I and 0 behaviors have a higher mean than the Type II and the Type II behaviors have a higher mean than the Type III.

Figure 2a. Protective Factors and Positive Behaviors across Clusters of Youth Aged 18 to 24 – Mexico

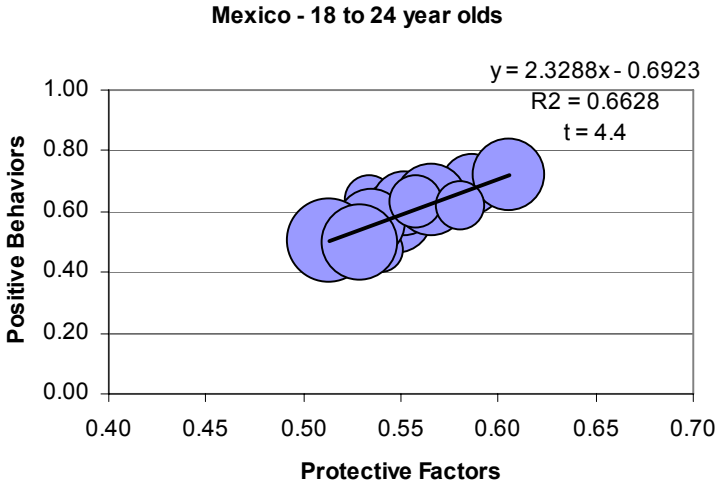


Figure 2b. Protective Factors and Positive Behaviors across Clusters of Youth Aged 18 to 24 – Chile

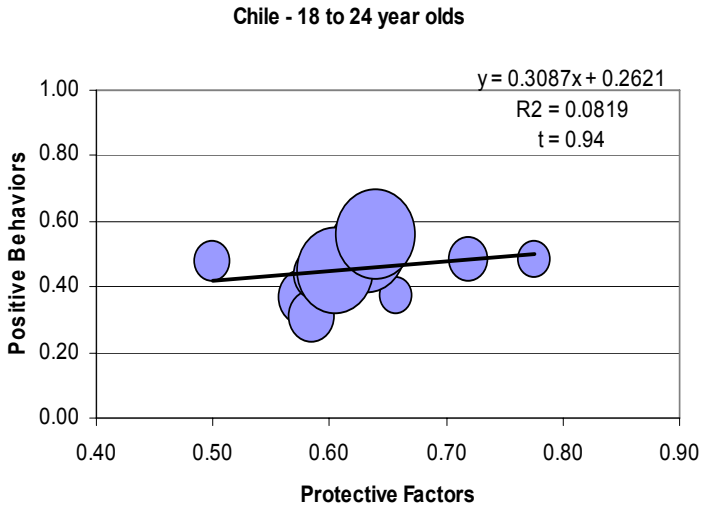


Figure 3a. Risk Factors and Positive Behaviors across Clusters of Youth Aged 18 to 24 – Mexico

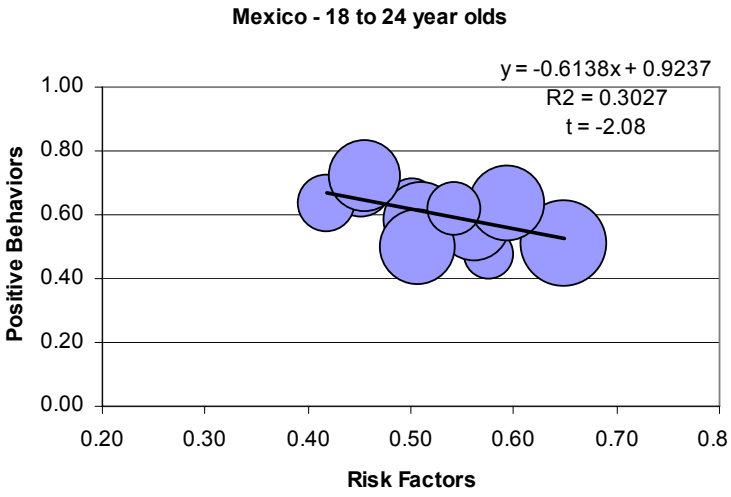


Figure 3b. Risk Factors and Positive Behaviors across Clusters of Youth Aged 18 to 24 – Chile

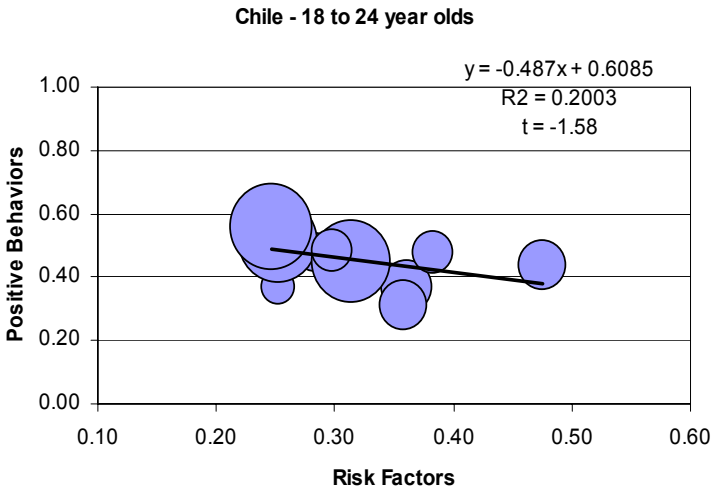


Figure 4a. Poverty and Positive Behaviors across Youth Aged 18 to 24 – Mexico

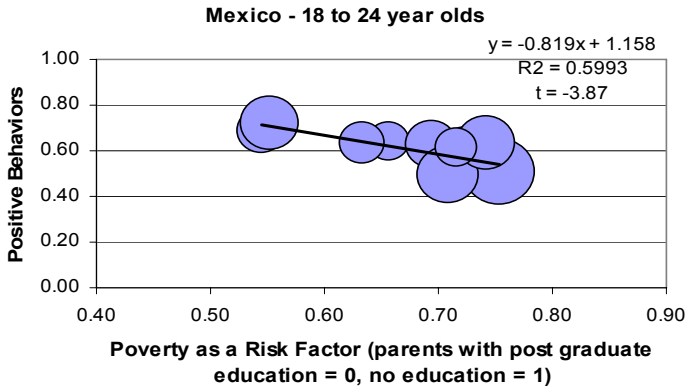


Figure 4b. Poverty and Positive Behaviors across Youth Aged 18 to 24 – Chile

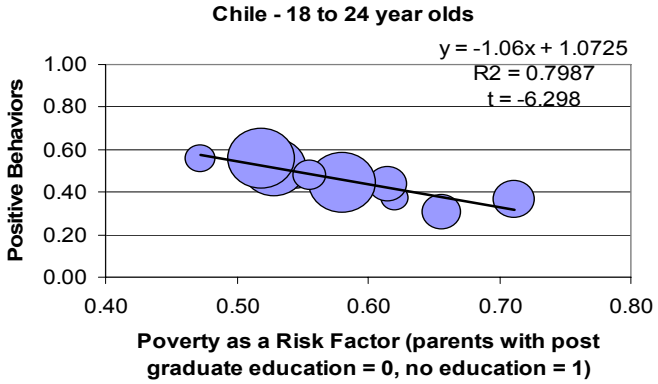


Figure 5a. Rural and Positive Behaviors across Youth Aged 18 to 24 – Chile

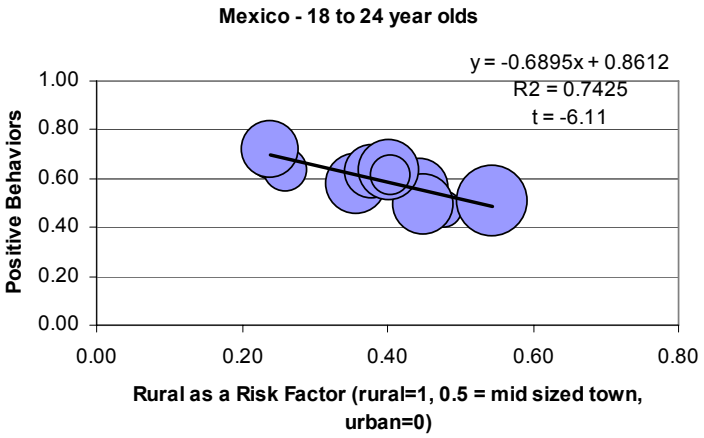


Figure 5b. Rural and Positive Behaviors across Youth Aged 18 to 24 – Chile

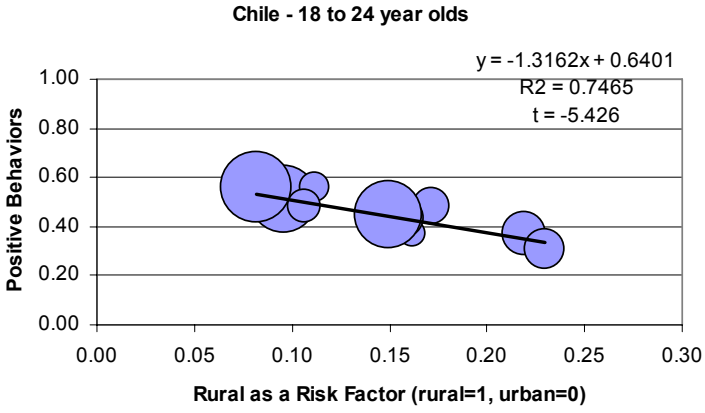
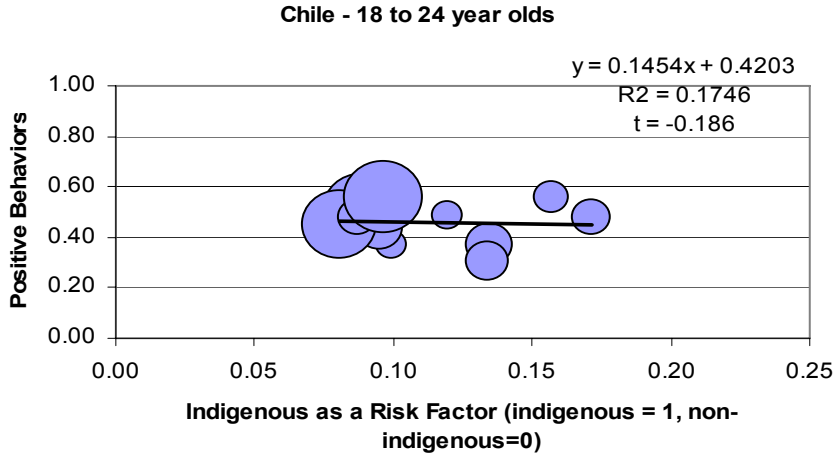


Figure 6. Chile Indigenous and Positive Behaviors across Youth aged 18 to 24



References

Attanasio, Orazio, Erich Battistin, Emla Fitzsimons, Alice Mesnard and Marcos Vera-Hernández, 2005, "How effective are conditional cash transfers? Evidence from Colombia" *Institute for Fiscal Studies Briefing Note No. 54*.

Becker, Gary and Kevin J. Murphy, 1988, "Theory of Rational Addition", *Journal of Political Economy*, 96, 675-700.

Bernat, Debra and Michael Resnick, 2006, "Healthy Youth Development: Science and Strategies", *Journal of Public Health Management Practice*, November (Supplement), S10-S16.

Blakely, Tony, Simon Hales, Charlotte Kieft, Nick Wilson, & Alistair Woodward, 2005, "The Global distribution of risk factors by poverty level", *Bulletin of the World Health Organization*, February, 83 (2).

Blum, Robert, 1998, "Healthy Youth Development as a Model for Youth Health Promotion", *Journal of Adolescent Health*, 22, 368-375.

_____, R. (2002) "Adolescent Health in the Caribbean." Unpublished paper, World Bank, Washington DC.

_____, R. and Marjorie Ireland, 2004, "Reducing Risk, Increasing Protective Factors: Findings from the Caribbean Youth Health Survey." *Journal of Adolescent Health* 35, 493-500.

_____, Marjorie Ireland and Sally-Ann Ohene, 2005, **"**The Clustering of Risk Behaviors Among Caribbean Youth**",** *Maternal and Child Health Journal*, 9(1), 91 - 100.

Bronfenbrenner, U., 1979, *The ecology of human development*. (Harvard University Press: Cambridge, MA).

Brook, Judith S., David W. Brook, Mario de la Rosa, Martin Whiteman, Erica Johnson, and Ivan Montoya, 2001, "Adolescent Illegal Drug Use: the Impact of Personality, Family, and Environmental Factors" *Journal of Behavioral Medicine*, 24(2), 183-203.

Brook, David W., Judith S. Brook, Zohn Rosen, and Ivan Montoya, 2002a, "Correlates of Marijuana Use in Colombian Adolescents: A Focus on the Impact of the Ecological/Cultural Domain," *Journal of Adolescent Health*, 31, 286-298.

Brook, David W., Judith S. Brook, and Tine Pahl (2002b) **"**The Longitudinal Relationship Between Drug Use and Risky Sexual Behaviors Among Colombian Adolescents" *Archives of Pediatric and Adolescent Medicine*, 156(11), 1101-1107.

Cunningham, Wendy and William Maloney, 2001, "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis", *Economic Development and Cultural Change*, 131-156.

Duda, R. and P. Hart, 1973, *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, NY.

Everitt, Brian, Sabine Landau and Morven Leese, 2001, Cluster Analysis, 4rd Edition, London: E. Arnold ; New York : Oxford University Press.

Gruber, Jonathan, Editor, 2001, Risky Behavior Among Youths: An Economic Analysis, University of Chicago Press.

Hall, Gillette and Harry Patrinos, 2006, *Indigenous Peoples, Poverty and Human Development in Latin America*, New York: Palgrave Macmillan.

Hutz, Claudio and Debora Rizzo M. Silva (2003). "Developmental Issues and Delinquent Behavior of Brazilian Adolescents," *International Society for the Study of Behavioural Development*, No. 2, Serial No. 42, pp 6-8.

Instituto Cidadania, 2004, *Projecto Juventude: Documento de Conclusão* (Brasilia, Brazil)

Lloyd, Cynthia B. editor, 2005, Growing up global: the changing transitions to adulthood in developing countries, Panel on Transitions to Adulthood in Developing Countries ; Committee on Population [and] Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Published: Washington, D.C.: National Academies Press.

Milligan, G.W. and M.C. Cooper. 1985. "An examination of procedures for determining the number of clusters in a data set." *Psychometrika*. 50(2), 159-179.

Resnick, Michael D., Marjorie Ireland, and Iris Borowsky, 2004, "Youth Violence Perpetration: What Protects? What Predicts? Findings from the National Longitudinal Study of Adolescent Health", *Journal of Adolescent Health*, 35, 424.e1–424.e10.

Scal, Peter, Marjorie Ireland, and Iris Borowsky, 2003, "Smoking Among American Adolescents: A Risk and Protective Factor Analysis" *Journal of Community Health*, 28(2).

World Bank, 2003, *Caribbean Youth Development: Issues and Policy Directions*. Washington, DC. World Bank.

_____, 2006, *World Development Report 2007: Development and the Next Generation*, Washington, DC. World Bank.

_____, 2007, *Youth At-Risk in Brazil*, Report No. 32310-BR, Washington, DC. World Bank.

Zweig, Janine M., Stacey D. Phillips, and Laura Duberstein Lindberg, 2002, "Predicting Adolescent Profiles of Risk: Looking Beyond Demographics", *Journal of Adolescent Health*; 31, 343–353.

Annexes

Annex A – Mean values of variables

	Mexico				Chile		
Variable	All ages	12 to 14	15 to 17	18 to 24	All ages	15 to 17	18 to 24
age	17.19	13.0	16.0	20.8	18.9	15.95	20.8
gender	0.54	0.51	0.53	0.56	0.53	0.51	0.55
Behaviors/Outcomes Average	0.62	0.56	0.64	0.61	0.51	0.59	0.47
older age when started working	0.54	0.70	0.49	0.46	0.65	0.76	0.58
not idle	0.80	0.91	0.82	0.70	0.73	0.93	0.64
no early school dropout	0.70	0.90	0.67	0.58	0.82	0.92	0.82
literacy	0.98	0.99	0.99	0.98	-	-	-
older age at onset of sexual activity	0.78	-	0.90	0.71	0.64	0.80	0.54
older age at first pregnancy/child	0.89	-	0.97	0.84	0.90	0.97	0.86
married	0.12	0.00	0.04	0.25	0.05	0.00	0.08
participate in activities	0.21	0.22	0.23	0.19	0.10	0.13	0.08
registered to vote	0.78	-	-	0.78	0.28	0.56	0.11
not victim	-	-	-	-	0.92	0.92	0.92
working	0.42	0.17	0.35	0.53	0.21	0.05	0.32
in school	0.45	0.86	0.62	0.27	0.60	0.91	0.41
years of education completed[a]	0.35	0.27	0.36	0.41	0.54	0.45	0.59
safe sex	0.61	-	0.90	0.59	0.52	0.78	0.35
has a child	0.19	-	0.03	0.29	0.19	0.04	0.28
low number of sexual partners in past year	0.94	-	0.98	0.92	-	-	-
attitude towards drugs	0.97	-	0.97	0.97	-	-	-
attitude towards alcohol	0.89	-	0.91	0.88	-	-	-
Protective Factors Average	0.57	0.58	0.58	0.55	0.63	0.64	0.62
connected	0.94	0.96	0.95	0.91	0.88	0.89	0.87
live with both parents	0.77	0.90	0.85	0.64	-	-	-
relationship with father	0.41	0.40	0.41	0.41	0.68	0.72	0.66
relationship with mother	0.50	0.48	0.50	0.51	0.85	0.88	0.83
connected to other adults	-	-	-	-	0.10	0.10	0.10
trust in governmental institutions	0.35	-	0.36	0.35	0.31	0.33	0.29
trust in community institutions	0.61	-	0.59	0.62	0.76	0.79	0.75
church attendance	0.38	0.42	0.38	0.35	-	-	-
school quality	0.80	0.81	0.80	0.79	-	-	-
optimism about future	-	-	-	-	0.88	0.88	0.87
preparation for future	-	-	-	-	0.84	0.79	0.87
communication with parents	0.48	0.51	0.46	0.46	-	-	-
spiritual influence / spirituality	0.19	0.17	0.19	0.20	0.33	0.36	0.31
sense of wellbeing	0.85	-	0.84	0.85	-	-	-
Risk Factors Average	0.56	0.57	0.56	0.56	0.28	0.25	0.30
parental influence (alcohol & smoking)	0.49	0.37	0.44	0.60	-	-	-
poor family cohesion	-	-	-	-	0.34	0.34	0.33
abuse in the home	-	-	-	-	0.06	0.06	0.06
substance abuse in the home	-	-	-	-	0.09	0.07	0.10
social exclusion	0.53	0.67	0.54	0.43	0.18	0.10	0.23
limited access to healthcare	0.49	0.50	0.51	0.48	-	-	-
community violence	-	-	-	-	0.25	0.26	0.25
parental response to misbehavior	0.28	0.29	0.26	0.29	-	-	-
parental response to good behavior	0.55	0.52	0.54	0.57	-	-	-
felt discriminated against	-	-	-	-	0.55	0.54	0.55

indigenous	-	-	-	-	0.11	0.11	0.10
live in a rural area	0.43	0.46	0.44	0.39	0.13	0.13	0.13
percentage living in rural area	0.25	0.28	0.26	0.22	0.13	0.13	0.13
low parental education[a]	0.69	0.69	0.69	0.69	0.57	0.58	0.57
household ownership of goods	0.78	0.79	0.78	0.77	-	-	-
monthly earnings household heads	0.80	0.80	0.80	0.81	-	-	-
low economic class (nse)	-	-	-	-	0.54	0.29	0.70
Sample size	37979	11871	9438	16670	5321	2057	3264

* highlighted rows are the variables that were treated as exogenous (they were not actively clustered on). *Notes*:
[a] In Mexico, a value of 0 indicates no education, 0.17 would be exactly 6 years of education, 0.34 would be exactly 9 years of education, 0.51 would be 12 years, 0.68 would be 15 years, 0.85 would be 19 or 20 years, and 1 would indicate a post graduate education. In Chile, a value of 0 indicates no education, 0.15 would represent from 1 to 7 years of education, 0.3 indicates a primary graduate (8 years completed), 0.45 from 9-11 years, 0.6 indicates a secondary graduate (12 years completed), 0.75 would represent from 13 to 16 years, 0.9 indicates a tertiary graduate, and 1 indicates postgraduate education.

Annex B – Description of the clusters

Table B1: Chilean Males 15-17

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
At-risk Risk type III 7.4% of the age-sex group (n=75)	Most idle Dropout Highest working (36%) Most with child (4%)	Lowest education Young start working Lowest participate in activities Early onset of sexual activity	Low connected (but not as low as the loners) Poor relationship with father	Rural (27%) Indigenous (20%) Feel socially excluded Most abuse in home Poorest 2nd most substance abuse in home
Risk type II Resilient 35.3% of the age-sex group (n=359)	In school Not idle 8% working	Have not dropped out Medium age started working Young sex 2nd most unsafe sex	Poorest relationship with father	Most abuse in home Most substance use in home
Risk type I Other connected so doing OK 8.6% of the age-sex group (n=88)	No kid 5% working Not idle Finished high school	Safer sex and start older Most vote	Most positive employment outlook Optimistic about future Trust community highly connected to others	More (but not most) rural No home abuse 2nd most community violence
Risk type I Loners 10.3% of the age-sex group (n= 105)	2nd most with child (3%) 2nd least in school (98%) Not idle 8% working (similar to "resilient" group)	Medium age start working and start sex 2nd lowest activities Lowest desire to vote (similar to "resilient" group)	Least optimistic about future Least prepared for future 2nd least spiritual Low trust in government 0% feel connected	Least rural 2nd least indigenous Medium poverty No home abuse No home drugs
Risk type 0 Parental connect & doing well 12.4% of the age-sex group (n =126)	No kids All in school 2% working	87% safe sex Older age start working & sex Higher participate in activities	Most optimistic about future Best prepared for future Most spiritual (by a lot) Most trust in community & government All connected Highest connect with mother and father	Least discrimination 2nd most indigenous 2nd least poor 0% socially excluded Best family cohesion No household abuse Most community violence
Risk type 0 Advantaged 26.0% of the age-sex group (n=265)	1% have kids All in school 1% working	Safest sex Oldest started working & sex Medium vote and participation	Highly connected 2nd most parental connection	Low discrimination 2nd least rural Least indigenous Wealthiest parents

				No household abuse High social exclusion

Table B2: Chilean females age 15-17

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III At-risk 9.1% of the age-sex group (n= 95)	Half with a child None in school 99% dropped out early 7% working Lowest educational attainment Most idle (93%)	Least safe sex Earliest start work & sex Least activities Most want to vote	Least optimistic about future Most prepared for future Lowest spirituality Most trust in government and 2nd most in community Poorer connect with parents, some connect with other adults	Most rural (25%) Poorest Most social exclusion Most abuse in household Some substance abuse in household Least community violence
Risk type II At-risk but connected 5.8% of the age-sex group (n= 60)	Few with children All in school Most working (93%) None idle	Second most unsafe sex Middle age start working and sex Most participate in activities	Least positive view on employment least trust in government or community all connected worst relationship with father or mother most connect to other adult	Most discrimination 2nd most rural (18%) Most indigenous (20%) Poorer Low social exclusion Worst family cohesion All abused in home Most home substance abuse
Risk type I Disconnected and socially excluded 12.8% of the age-sex group (n = 133)	5% have a kid All in school None working None idle	Many report safe sex (82%) Older when start working & sex	Most positive view on employment Least optimistic about future Least prepared for future Least trust in community Lowest connected (41%) Poor relationship with father, but OK with mother Not connected to other adults	Least rural (8%) Poorer 2nd most social exclusion Reports 1st most violent community
Risk type 0 Advantaged (connected with parents but not with other adults) 63.6% of the age-sex group (n = 661)	2% with kid All in school None working	Safest sex Oldest start work/sex	Most optimistic about future Most spiritual Most trust in community All connected Most connected to father, mother No connection to other adults	Low rural or indigenous no social exclusion best family cohesion no abuse in hh
Risk type 0 Advantaged (connected with all)	1% with kid None working All in school	Safer sex Oldest start work/sex	Most optimistic about future Most spiritual All connected	Low rural or indigenous No social exclusion No abuse in hh

8.7% of the age-sex group (n = 90)			2nd most connected to mother and father and all connected to other adults	

Table B3: Chilean males age 18-24

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III At-risk 16.0% of the age-sex group (n = 237)	Few in school (13%) 2nd most working 22% with kid Fewest finished secondary school (28%) Lowest education level	Least safe sex Youngest start work/sex Least married Least participate in activities	Least positive view on employment Least spiritual Least trust in government or community Least connected (75%) Worst relationship with father	Most rural Poorest Worst family cohesion Most substance abuse in home (42%)
Risk type III Idle graduates 7.5% of the age-sex group (n = 111)	None working None in school All idle All finished school High education level	Safest sex Oldest start work/sex Least married	Least optimistic about future All connected Best relationship with father	Least discrimination High rural Middle poverty All socially excluded
Risk type III Young adults 11.4% of the age-sex group (n = 169)	Most working (57%) 25% with kid 20% idle	Most married (8%) Least participate in activities 2nd highest vote	Most prepared for future	High rural Most indigenous Middle poverty None socially excluded
Risk type I Active and upbeat 7.9% of the age-sex group (n = 117)	½ in school 20% idle 15% early drop	Safe sex High participate in activities High vote	Most positive view on employment, future, and prep for future Most spiritual All connected Best relationship with mother All connected to other adults	Least rural
Risk type I Resilient 9.3% of the age-sex group (n = 137)	1/3 working ½ in school No secondary school drop	Safe sex	Least positive view on employment and least prep for future Low spirituality Least trust in government and community None connected Worst relationship with mother	Least rural Least indigenous Worst family cohesion Abuse in household (9%) Substance abuse in household (15%)

Risk type 0 Advantaged 38.9% of the age-sex group (n = 576)	40% working Most in school (64%) Second most education Most not idle No secondary school dropout	Safe sex Older working & sex	All connected Best connected with mother	Least rural Least indigenous Low poor Abuse in household (8%) Socially excluded
Risk type 0 Advantaged voters 9.0% of the age-sex group (n = 134)	40% working Most education 20% idle Lowest number with children	Highest vote Older working and sex	Optimistic about the future, employment, etc… High spirituality Highest trust in the government Medium connected	Highest discrimination High indigenous (16%) Least poor Socially excluded

Table B4: Chilean females age 18-24

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III at-risk 11.7% of the age-sex group (n = 209)	12% working None in school 88% with a kid 88% idle All secondary school dropout Low education level	Least safe sex Earliest work & sex Most married (30%) 6% report being a victim	Least optimistic about future, or prep for future Least spiritual Most trust in government Worst relationship with mother Second worst relationship with father	Most rural (23%) Most indigenous (13%) Poorest Most socially excluded (58%) Most community violence
Risk type II resilient 11.8% of the age-sex group (n = 210)	¼ working 38% in school ½ with a kid ½ idle 73% finished school	safe sex Medium age start working & sex 5% married	Least optimistic about future Least trust in government Worst relationship with father and mother Some connected to other adult (13%)	Most discrimination Medium rural (16%) Medium poverty Medium social exclusion Worst family cohesion ½ abused in home 2/3 substance abuse in home
Risk type II young adults (start behaviors later in life but outcomes earlier) 32.6% of the age-sex group (n = 582)	¼ working 16% in school ½ with a kid 1/3 idle Most finished secondary school	safe sex 18% married	Most positive view of employment Most spiritual Least connected (overall)	Medium rural Medium poverty Medium social exclusion
Risk type II Connected 9.0% of the age-sex group (n= 160)	Most working (33%)	Most participate in activities	Least positive view on employment Best prepared for future All connected	Low rural Medium poverty No abuse in home

			All connect with other adults	
Risk type 0 Advantaged 34.9% of the age-sex group (n = 622)	Most in school (75%) Most educated Least with a kid (15%) Not idle (7% are idle) No secondary school dropout	Safest sex (53%) Oldest when start work & sex Least married (1%) Most participate Most vote	Most optimistic about future All connected Best relationship with mother and father	Least rural Least poor Least socially excluded No abuse in home

Table B6: Mexican males age 12-14

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III at-risk 13.0% of the age-sex group (n=758)	Half working None in school Half idle Lowest literacy (93%) Least participate in activities	Medium age start work	Fewest live with both parents (87%) Worst relationship with father and mother Lowest church attendance Worst school quality	Some social exclusion (58%) Most limited access to healthcare Most parental drug influence Most parental response to good and bad behavior Most rural Poorest
Risk type II connected but early workers 20.5% of the age-sex group (n= 1193)	42% working All in school None idle All literate	Youngest age start working Most participate in activities	Least connected Best relationship with mother and father Most church attendance Most spiritual influence	58% socially excluded Medium poor
Risk type I loners (advantaged, but loners) 32.5% of the age-sex group (n= 1893)	None working All in school None idle All literate	Never worked	all connected Most live with both parents (93%) Better relationship with parents Best communication with parents Most go to church with parents Best school quality	All socially excluded Most access to healthcare Least parental influence drugs Least parental response to misbehavior Least poor
Risk type 0 advantaged 34.1% of the age-sex group (n= 1989)	20% working All in school None idle All literate	Medium age start working	All connected Better relationship with parents Best school quality	None socially excluded Least rural Least poor

Table B7: Mexican females age 12-14

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III At-risk 15.2% of the age-sex group (n = 908)	Most working (27%) Most idle (72%) Most illiterate (8%) Least participate in activities	None in school Lowest activities	Fewest live with both parents (14%) Worst relationship with father and mother Worst communication with parents Fewest activities with parents Worst school quality	73% socially excluded Most limited access to healthcare Most parental response to good behavior Most rural Poorest
Risk type II Early workers 45.5% of the age-sex group (n = 2723)	Some working (13%) None idle All literate	All in school Medium age start working	Better relationship with mother Medium on all the variables	All socially excluded Most limited access to healthcare Most parental response to drugs Medium poor
Risk type I Advantaged loners 16.4% of the age-sex group (n= 982)	17% working All literate None idle Oldest work age	All in school	Most live with both parents (93%) Best relationship with father and mother Best communications with and activities with parents	All socially excluded All access to healthcare Least parental influence drugs Most parental response to good behavior Least rural Least poor
Risk type 0 Advantaged youth 23.0% of the age-sex group (n= 1376)	11% working All literate None idle	All in school Medium early work Worst attitude towards drugs and alcohol	Best relationship with parents	None socially excluded More limited access to healthcare Second least poor

Table B7: Mexican males age 15-17

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III At-risk 39.8% of the age-sex group (n= 1765)	Most working (70%) 25% idle 3% illiterate 6% in school Lowest education level	Almost none in school Least safe sex (15%) Youngest when started working & sex Least participate in activities	Least connected (but still 91%) Worst relationship with father and mother Worst communication and activities with parents Worst school quality Worst trust in government Lowest sense of well-being	Half socially excluded Most parental influence drugs Most parental response to mis- and good behavior Most rural Poorest
Risk type II High risk	37% working All in school	Most (93%) safe sex Medium age started working	Connected with better relationship with father	Most limited access to healthcare

25.3% of the age-sex group (n=1120)	None idle All literate	Latest start sex	Highest going to church with parents Lowest trust in government	Half socially excluded Higher parental response to behavior More poor
Risk type I Loners (rename?) 15.4% of the age-sex group (n= 681)	27% working All in school None idle All literate	Safer sex (90%) Oldest age start working Older age start sex	Best relationship with father and mother Most activities and communication with parents Highest sense of well-being	All socially excluded All access to healthcare Least parental influence drugs Least parental response to behaviors
Risk type 0 Advantaged 19.6% of the age-sex group (n= 869)	27% working All in school None idle All literate	Safer sex (90%) Older age start working Older age start sex	Higher relationship with father and mother Most trust in government	None socially excluded All access to healthcare Least poor Least rural

Table B8: Mexican females age 15-17

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III Idle early dropouts 18.7% of the age-sex group (n= 932)	Idle Not parents Lowest literacy (95%) All working None in school Lowest education level All early school drop	Medium age when started working & sex Medium level of activity participation (15%)	Well connected Most live with both parents Second lowest relationships with parents Second lowest church attendance Worst school quality Worst sense of wellbeing	Socially excluded (68%) Second lowest access to healthcare Most rural Most poor
Risk type III Early workers 12.1% of the age-sex group (n= 603)	None Idle All working None in school All dropped out early Second lowest education level None are parents	Earliest age when started working Medium/safe sexual behavior Medium activities	Connected Most live with both parents Medium relationships with parents Medium spiritual influence Second lowest sense of wellbeing	High social exclusion High parental influence with drugs and alcohol Medium rural Second poorest
Risk type III Young wives 7.3% of the age-sex group (n= 363)	17% working 6% in school Second lowest education level Second lowest literacy (96%) 80% idle 54% are mothers 66% married	Earliest sexual activity Most having sex and most unprotected and most with more than 1 partner in the last year Medium early work Lowest involvement in activities	Second least connected Fewest live with 1 or both parents Worst relationships with parents Lowest spiritual influence & church attendance Second lowest school quality	Not very socially excluded Low access to healthcare Worst parental influence, response to misbehavior and to good behavior Middle rural Poor

			Lowest trust in government and community institutions Medium sense of well being (85% report being happy)	
Risk type I Resilient loners 19.2% of the age-sex group (n= 957)	Not idle 23% working 95% in school Not parents, fully literate	High activities Medium age at first job Best sexual behavior	Lowest connection Second best relationships with parents Highest spiritual influence & church attendance Highest school quality Medium/low sense of well being (82% report being happy)	No social exclusion Worst access to healthcare Medium rural Medium poor
Risk type I Advantaged loners 19.0% of the age-sex group (n= 947)	Not idle 14% working 95% in school Only 4% early dropout Not parents, fully literate	High activities Oldest age at first job Good sexual behavior	Highest live with both parents Best relationships with parents Highest school quality Highest sense of wellbeing	No social exclusion Second lowest rural Least poor Best parental response to good behavior/misbehavior and influence re alcohol and drugs
Risk type 0 Advantaged youth 23.7% of the age-sex group (n= 1178)	Not idle 17% working Not parents, fully literate	Highest activities Medium age at first job Good sexual behavior Worst attitude towards alcohol (11% can justify getting drunk)	Second best relationships with parents Medium school quality Second highest sense of wellbeing	All socially excluded Least rural Least poor Second best parental response to good behavior/misbehavior and influence re alcohol and drugs

Table B9: Mexican males age 18-24

Cluster	Outcomes	Behaviors	Protective Factors	Risk Factors
Risk type III Idle 10.1% of the age-sex group (n = 743)	All idle Low education level 73% dropped out early 99% literate	Medium sex Low activities Second worst attitude towards drugs 14% married	Good connection Second worst relationship with father Worst relationship with mother Worst trust in institutions (both government and community) Lowest sense of wellbeing (82%)	Some social exclusion (37%) Worst access to healthcare Parental influence towards alcohol and smoking (low variable → high involvement) Most rural (31%) Poorest
Risk type III Drinkers 20.6% of the age-sex group	Lowest literacy (90%) 50% dropped out early Not idle	Medium sex Youngest age at first job 14% married	Good connection Worst relationship with parents Lowest church attendance	Some social exclusion (36%) Second worst access to healthcare

(n= 1511)	80% working 30% in school Lowest education level		Worst school quality Second lowest sense of wellbeing	Low parental influence Second most rural (27%) Second poorest
Risk type III Working dads 11.4% of the age-sex group (n= 837)	All working 7% in school Low education level achieved 71% have a child	99% married Most unsafe sex Worst number of sexual partners Low participation in activities Best attitudes to drugs & alcohol	Good connection Most live with neither parent Most go to church with parents Highest sense of wellbeing	Limited social exclusion (10%) Best access to healthcare Medium parent influence Worst parental response to poor behavior 20% rural Second highest parental education Low luxury goods
Risk type II Early dropouts 24.2% of the age-sex group (n= 1778)	All early school dropout 91% working Low education level Not idle	Second worst attitude towards drugs Early work Second earliest sex Worst number of sexual partners Not married (5% married) Medium activities	Connected Medium relationships with parents Lowest church attendance with parents Lowest spiritual influence Medium school quality Low trust in institutions Medium sense of wellbeing	No social exclusion Second/third poorest Highest parental influence Highest parental response to good behavior, second highest to bad behavior
Risk type I Resilient loners 18.2% of the age-sex group (n= 1335)	Not idle 78% working 28% in school 75% early drop Literate Medium education level 11% have children	Second safest sex Medium age at start work Oldest age at onset of sex Best number of sexual partners Low level of activities	Lowest level of connectedness Most live with 1 or both parents Good relationships with father and mother High spiritual influence High school quality High trust in government and community Medium wellbeing	Socially excluded Second highest parental influence Medium poor Second best access to healthcare
Risk type 0 Advantaged students 15.5% of the age-sex group n= 1142	Very high education level 74% in school 56% working None idle 3% have children	Worst attitude towards drugs (77%) High level of activities Oldest age at first job Second oldest at start work Safest sex	Connected Most live with 1 or both parents Best relationships with parents High church attendance High school quality High trust in community	Third highest parental influence Least poor (by far)

			Medium trust in government High wellbeing	

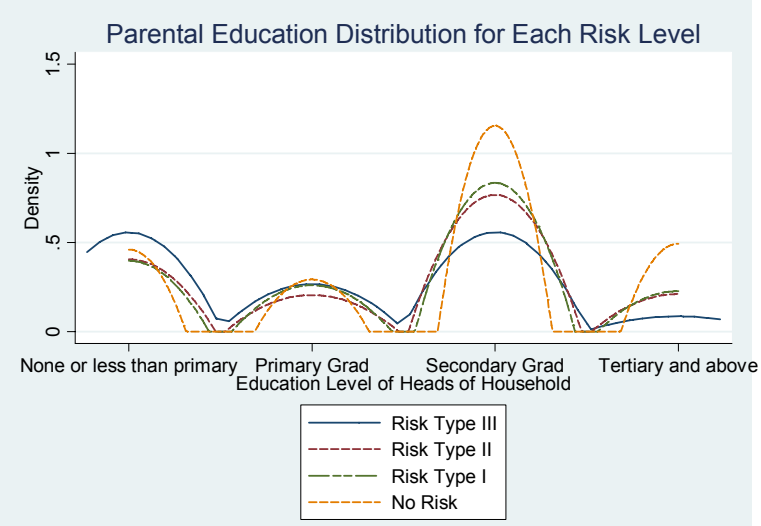
Table B10: Mexican males age 18-24

Cluster	**Outcomes**	**Behaviors**	**Protective Factors**	**Risk Factors**
Risk type III At-risk 19.4% of the age-sex group (n = 1810)	All idle 93% early dropout Lowest education level (even when compared to males) 68% have children 63% married	Low safe sex Earliest age at first sex Most partners in past year Younger age at first pregnancy Lowest activities Second worst attitude towards alcohol	Few live with one or both parents Connected Second poorest relationships with parents Lowest spiritual influence Second highest level of happiness	Not socially excluded Half have access to healthcare Medium rural (26%) Medium poor Highest level of parental attempted control over behaviors
Risk type III At-risk loners 24.9% of the age-sex group (n= 2324)	93% idle (5% working, 2% in school) 81% early dropout Lowest education level Almost all illiterates in this group (8% illiterate) 40% married	Average safe sex Early sex Lowest activities Worst attitude towards alcohol	Lowest connected Medium "live with parents" Worst relationships with parents Medium church/spiritual influence Lowest school quality Low happiness	Socially excluded Lowest access to healthcare Most rural (36%) Poorest
Risk type II Working dropouts 18.2% of the age-sex group (n=1693)	All early dropout Not idle 83% working, 29% school 23% with child Medium education level	Second earliest work Medium sex Medium activities	Medium connected Medium/high live with parents	High social exclusion Poor access to healthcare Low rural
Risk type II Early workers doing ok 10.7% of the age-sex group (n= 999)	High early dropout (90%) 40% with child Medium education level 45% married	Earliest work Medium/unsafe sex	Connected Medium live with parents Medium relationships with parents Highest happiness	Best access to healthcare Second least poor
Risk type I Resilient poor 9.1% of the age-sex group (n= 850)	Not idle 60% working, 48% in school Medium education level	Safest sex Worst attitude towards alcohol	Connected Highest live with parents Medium relationships with parents Highest church attendance	Medium rural Medium/high poor
Risk type 0 Advantaged students 17.6% of the age-sex group (N= 1644)	Not idle 40% working, 69% in school Highest education level (very high)	Safest sex Worst attitude towards drug use	Highest live with parents Best relationships with parents Highest spiritual influence High sense of wellbeing	Most independent from parents (alcohol and smoking) Lowest rural Least poor

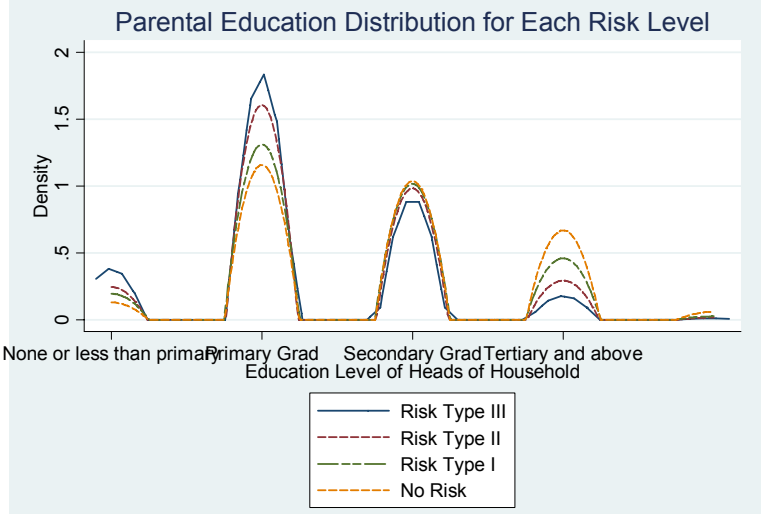
Annex C – Kernel density graphs of parental education level by risk level

The figures[38] below demonstrate that the more at-risk clusters in both Chile and Mexico have a lower level of parental education than the other clusters. The education level of parents or heads of household = 0 if no education or less than a primary school education, 1 if primary graduate, 2 if secondary graduate, 3 if tertiary graduate and 4 if above tertiary level. In Chile, parental education level is more strongly related than the assigned socio-economic level (nse)[39]. This could be due to unobserved characteristics related to educational achievement. We also see that Chile has a higher parental education level than Mexico.

Chile Kernel Density diagram of parental education level by risk level



Mexico Kernel Density diagram of parental education level by risk level



[38] Generated using kernel density plots with Epanechnikov kernel.
[39] Details available upon request